

DOCUMENT RESUME

ED 183 525

SP 015 642

AUTHOR Fisher, Charles W.; And Others
TITLE Teaching Behaviors, Academic Learning Time and Student Achievement: Final Report of Phase III-B, Beginning Teachers Evaluation Study. Technical Report V-1.
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.
SPONS AGENCY California State Commission for Teacher Preparation and Licensing, Sacramento.; National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jun 78
CONTRACT 400-75-0001
NOTE 493p.
AVAILABLE FROM Commission for Teacher Preparation and Licensing, 1020 "O" St., Sacramento, CA 95814 (\$25.00).
EDRS PRICE MF02/PC20 Plus Postage.
DESCRIPTORS *Academic Achievement; Educational Research; Effective Teaching; *Interaction Process Analysis; Learning Processes; Research Methodology; Student Behavior; *Teacher Behavior; Teacher Evaluation; *Teaching Methods; *Time Factors (Learning)

ABSTRACT

Four major questions are addressed in this research report: What is the relationship between student academic learning time and student achievement?: What are the relationships between teaching processes and academic learning time?: Are teaching processes, academic learning time, and student achievement related to student attitudes?: and, Are instructional variables related to retention of achievement over the summer? An Academic Learning Time Model of classroom instruction is presented. Design, instrumentation, and data collection methods of this research project are outlined. Analyses of the joint relationships among the variables--teaching processes, academic learning time, and student achievement--are presented, as well as analyses of student attitudes and retention of achievement. (JD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED183525

PRE PUBLICATION COPY
This document is still
subject to review and revision

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED
EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING
IT. POINTS OF VIEW OR OPINIONS STATED
DO NOT NECESSARILY REPRESENT
THE OFFICIAL POSITION OR POLICY OF
THE NATIONAL INSTITUTE OF EDUCATION.

THIS DOCUMENT REPRODUCES THIS
DOCUMENT HAS BEEN REPRODUCED BY
Kenneth W. Arms
EDUCATIONAL RESEARCH
INFORMATION CENTER

BTES

BEGINNING TEACHER
EVALUATION STUDY

Technical Report Series

Technical Report V-1

Teaching Behaviors, Academic Learning
Time and Student Achievement: Final
Report of Phase III-B, Beginning
Teacher Evaluation Study

Charles W. Fisher
Nikola M. Filby
Richard Marlette
Leonard S. Cohen
Marilyn M. Dshaw
Jeffrey E. Moore
David C. Berliner

June, 1978



FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT

100 POLARIS STREET • SAN FRANCISCO, CALIFORNIA 94109 • (415) 774-1000

This work was completed pursuant to Contract #400-75-0001 between the Department of Health, Education and Welfare, National Institute of Education and the California Commission for Teacher Preparation and Licensing.

The Far West Laboratory for Educational Research and Development, 1855 Folsom Street, San Francisco, California 94102, is a nonprofit organization supported in part by the United States Office of Education and the National Institute of Education, Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

Technical Report V-1

Teaching Behaviors, Academic Learning Time
and Student Achievement: Final Report
of Phase III-B,
Beginning Teacher Evaluation Study

Charles W. Fisher
Nikola N. Filby
Richard Marliave
Leonard S. Cahen
Marilyn M. Dishaw
Jeffrey E. Moore
David C. Berliner

June, 1978

FEB 13 1980

PREFACE

The study reported here is the culmination of a multi-year research program on teaching effectiveness. Phase III of the Beginning Teacher Evaluation Study (BTES) was conducted by a research team at the Far West Laboratory for Educational Research and Development. The work was carried out between 1974 and 1978 while under contract to the California Commission for Teacher Preparation and Licensing (CCTPL) with funds provided by the National Institute of Education. The CCTPL initiated this research program in 1972 in the belief that sponsorship and encouragement of basic research on teaching and learning will lead to improvements in teacher training and therefore, will enhance the day-to-day educational experiences of children. The CCTPL, which has a large number of responsibilities concerning the training and licensure of teachers in California, has maintained its commitment to this belief. The BTES has been supervised by the Research and Planning Committee of the CCTPL. Throughout Phase III, the Committee members and CCTPL staff have contributed substantially to the study.

Approximately twice each year, during the course of the study, the CCTPL brought together a 25 member Research Advisory Board representing teachers, teacher educators, administrators, subject matter specialists, and educational researchers to help guide the conduct of the study. At various stages of the study, the Far West Laboratory research team made presentations to the Research Advisory Board and received very helpful feedback on research designs and interim reports. During the final year of the study, a subgroup of the Research Advisory Board was set up by the CCTPL to provide continuing feedback on the analyses documented in this report. This group, consisting of Drs. Leigh Burstein, Walter Borg, Gary Fenstermacher,

Marjorie Powell, Tom Romberg, Barak Rosenshine, and Rod Skager, were extremely knowledgeable about the details of the study and provided very insightful and valuable guidance on a number of substantive and methodological points.

Throughout the Project, Dr. Calvin Wright of Educational Research and Evaluation, Incorporated, monitored our research program. The objectivity and insights provided by Dr. Wright and his staff substantially improved the conduct of the study.

Phase III of BTES was designed and carried out by a team of researchers. Since the study was a relatively large undertaking, each person filled a number of roles over the four-year period. Phase III-A, Phase III-A Continuation, and the first portion of Phase III-B were directed by Dr. David Berliner. Phase III-A was especially strongly influenced by him and it was this foundation which made the later work possible. In 1977, Dr. Berliner assumed the position of Head, Department of Educational Psychology, University of Arizona. While he was no longer based in San Francisco, Dr. Berliner maintained an active role in the research program.

The final year of the project, including major data analyses and preparation of reports, was directed by Dr. Charles Fisher. Throughout the study, individual members of the staff took major responsibility for specific areas of analysis and writing. For this report, analysis and writing on teaching behaviors was headed by Dr. Nikola Filby and work on the relationship of academic learning time and student achievement was led by Dr. Richard Marliave. The analysis of outlier classes was headed by Dr. Leonard Cahen. Collection and analysis of teacher records of allocated time were directed by Marilyn Dishaw. Jeffrey Moore directed the data processing group.

We would like to express our appreciation to the many persons who

contributed directly to the completion of the Phase III-B study.

Pearl Howell and Carol Rice coordinated all field activities during the data collection phase of the study. The data were collected by a team of 21 field workers. We appreciate the long hours of effort they put into administering tests, interviewing teachers, and completing the extensive rounds of observation.

Literally millions of pieces of data were collected during Phase III-B. The gargantuan job of preparing the data for processing, editing, and carrying out computer analyses was done by Jeffrey Moore, Fannie Walton and Patricia Storm. Their efficiency, accuracy, and good humor are very much appreciated.

Throughout the study and especially during the preparation of technical notes and reports, large amounts of typing and clerical support were required. These services were coordinated and competently carried out by Edna Rohnett and Deborah Walton, with "last month" assistance from Vicki Tiernan.

As our work progressed we benefited greatly from the writings of and interchanges with a large number of colleagues. We especially appreciate the contributions of Annegret Harnischfeger and David Wiley, who started us thinking about instructional time; and Robert Linn and Richard Shavelson who provided sound advice in a number of technical as well as substantive areas.

We also acknowledge the assistance of the many people not directly involved with the programmatic side of the research effort. Without their help, it is unlikely that much would have been accomplished in the research domain. For continued institutional support, our appreciation goes to Dr. John Hemphill, Director of the Far West Laboratory for Educational Research and Development and to Dr. Beatrice Ward, Deputy Director of the Laboratory. Special thanks goes to K. H. Harms, Deputy Associate Laboratory Director, for frequent advice and strong support on fiscal management.

We also acknowledge the continued interest and support of this project by Drs. Virginia Koehler and Joseph Vaughan of the National Institute of Education.

Finally, we give special recognition to the teachers, administrators and students in California schools who, by volunteer participation, made this research possible. We thank the following school districts for their cooperation: Alameda, Fremont, New Haven, Oakland, Richmond, San Bruno, San Francisco, San Lorenzo, San Mateo, and Vallejo. The principals, teachers, and students from these districts responded to the time-consuming activities that field research requires.

The teachers participating in this study worked very hard to provide us with accurate information about their classes. We appreciate their openness in allowing us to observe them and their students over extended periods of time.

Charles W. Fisher
Nikola N. Filby
Richard Marliave
Leonard S. Cahen
Marilyn M. Dishaw
Jeffrey E. Moore
David C. Berliner

TABLE OF CONTENTS

PREFACE	i
LIST OF TABLES	x
LIST OF FIGURES	xv
CHAPTER 1: INTRODUCTION	1-1
Context for Phase III-B	1-3
Structure of the Report	1-8
CHAPTER 2: THE ACADEMIC LEARNING TIME MODEL OF CLASSROOM INSTRUCTION . .	2-1
Introduction	2-1
Instructional Processes	2-8
Classroom Environment	2-11
Summary	2-13
CHAPTER 3: OVERVIEW OF DESIGN AND INSTRUMENTATION	3-1
Design of the Correlational Field Study	3-1
Sample	3-2
Student Achievement Measures	3-6
Student Attitude Variables	3-8
Academic Learning Time Variables	3-8
Allocated time from teacher logs	3-8
Allocated time, engagement rates and error rates from observation	3-9
Comparison of allocated time from teacher logs and direct observation system	3-12
Measures of allocated and engaged time	3-14
Comparison of sources of time data	3-16
Student error rate	3-17
Teaching Process Measures	3-20
Interactive teaching behavior	3-20
Measures of planning and general characteristics	3-21
Initial measures -- interviews	3-22

Initial measures -- daily rating scales	3-24
Initial measures -- aptitude and attitude	3-25
Process of variable selection and definition	3-26
CHAPTER 4: ACADEMIC LEARNING TIME AND STUDENT ACHIEVEMENT	4-1
The ALT Variables	4-2
Descriptive Data on ALT	4-6
General time allocations	4-6
Specific time allocations	4-6
Intercorrelations of ALT variables	4-12
Relationships between ALT and Student Learning	4-15
A-B analyses	4-16
B-C analyses	4-23
Discussion: ALT and Student Learning	4-35
Non-experimental	4-35
Strength of the findings	4-36
Generalizability	4-39
Follow-up analyses of low error rate	4-41
Summary: ALT and Student Learning	4-50
CHAPTER 5: TEACHING PROCESS VARIABLES AND STUDENT ENGAGEMENT	5-1
INTERACTIVE TEACHING BEHAVIORS AND STUDENT ENGAGEMENT	5-1
The Interactive Teaching Behaviors	5-2
Correlations with Engagement	5-10
Teaching Behavior, Setting, and Engagement	5-15
GENERAL CLASSROOM VARIABLES AND STUDENT ENGAGEMENT	5-24
The General Variables	5-24
Residual variables	5-29
Descriptive Statistics	5-31
Correlations with Engagement	5-35
CHAPTER 6: TEACHING PROCESS VARIABLES AND STUDENT ERROR RATE	6-1
DIAGNOSIS AND PRESCRIPTION AND STUDENT ERROR RATE	6-3

The Diagnosis and Prescription Variables	6-3
Prescription	6-5
Descriptive Statistics	6-7
Correlations with Low Error Rate and High Error Rate	6-10
INTERACTIVE TEACHING BEHAVIORS AND STUDENT ERROR RATE	6-20
Summary	6-25
CHAPTER 7: TEACHING PROCESS VARIABLES AND ACADEMIC LEARNING TIME AS PREDICTORS OF STUDENT ACHIEVEMENT	7-1
Correlations Between Teaching Process Variables and Student Achievement	7-1
Differences between achievement subtests	7-25
Regressions Using ALT Variables and Teaching Process Variables to Predict Achievement.	7-27
Summary	7-34
CHAPTER 8: A QUANTITATIVE AND QUALITATIVE EXAMINATION OF SELECTED SECOND GRADE CLASSES	8-1
Purpose of this Form of Inquiry	8-2
Methodology	8-4
Residualized achievement measures	8-5
Outliers	8-7
Quantitative Data on Selected Classes	8-9
Descriptions of a Set of Outlier Classes	8-19
Class L (Category I)	8-19
Class Q (Category I)	8-20
Class V (Category II)	8-25
Class F (Category II)	8-26
Class H (Category III)	8-27
Class O (Category III)	8-29
Discussion	8-32
Summary	8-36

CHAPTER 9: STUDENT ATTITUDES AND SCHOOL INSTRUCTION	9-1
Description of the Attitude Data	9-2
Student Attitude and Achievement	9-6
Student Attitude and Academic Learning Time	9-9
Student Attitude and Teaching Process Measures	9-13
Summary	9-19
CHAPTER 10: SUMMER RETENTION OF ACHIEVEMENT	10-1
Introduction	10-1
Sample	10-2
Pattern of Achievement Scores A-D	10-4
Retention of Achievement C-D	10-8
Relationship between Instructional Variables and Achievement Retention	10-16
Summary	10-19
CHAPTER 11: SUMMARY AND CONCLUSIONS	11-1
What is a "Conclusion?"	11-2
Limitations to the Study	11-5
Conclusions	11-7
Teaching process variables	11-12
What next?	11-25
Diagnosis	11-26
Prescription	11-28
Presentation	11-32
Monitoring	11-33
Feedback	11-34
Context	11-38
A Last Word	11-41
REFERENCES	R-1
REFERENCE NOTES	R-2
Appendix 1-A: Annotated Bibliography of Technical Notes and Reports: Phase III Beginning Teacher Evaluation Study	1A-1

Appendix 3-A:	Descriptive Information on Teachers and Target Students	3A-1
Appendix 3-B:	Descriptive Information on Socioeconomic Status for Target and Non-Target Students	3B-1
Appendix 3-C:	Psychometric Data on Achievement Tests and Attitude Scales	3C-1
Appendix 3-D:	Teacher Log Content Categories	3D-1
Appendix 3-E:	Brief Description of Observation Categories	3E-1
Appendix 3-F:	Inter-observer Reliability Coefficients for Observation Categories	3F-1
Appendix 3-G:	Correspondence among Observation, Congruency, and Teacher Log Content Categories	3G-1
Appendix 3-H:	Reliability of Observed Engagement Variables	3H-1
Appendix 3-I:	Reliability of Observed Student Error Rates	3I-1
Appendix 4-A:	Non-Instructional Activities	4A-1
Appendix 7-A:	Correlations of All Teaching Process Measures with Residual Achievement (A-C), by Subtest	7A-1
Appendix 7-B:	Intercorrelation of Teaching Process Measures	7B-1
Appendix 10-A:	Regression of C Scores on A Scores, Academic Learning Time and Teaching Process Variables on the Subsample of Students Who Were Tested at D	10A-1

LIST OF TABLES

<u>Table</u>	<u>Page</u>
4-1 Where the Day Goes --Minutes per Day in General Content Categories	4-7
4-2 Descriptive Data, Academic Learning Time, Time Allocations in Specific Content Categories (A-C Period), Reading and Mathematics, Grade 2 (N=139)	4-8
4-3 Descriptive Data, Academic Learning Time, Time Allocations in Specific Content Categories (A-C Period), Reading and Mathematics, Grade 5 (N=122)	4-9
4-4 Descriptive Data, Academic Learning Time, Engagement and Error Rate (A-C Period)	4-11
4-5 ALT Variable Intercorrelations, Allocated Time (Total Reading and Total Mathematics for the AB and BC Periods) Period), Low Error Rate (A-C)	
4-6 Regressions Analyzing the Effects of Engaged Time, Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the A-B Period, Regressing Post Achievement on Pre-Achievement and Engaged Time	4-17
4-7 Regressions Analyzing the Effects of Low and High Error Rates, Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the A-B Period, Regressing Post Achievement on Pre Achievement, Engaged Time, Low Error Rate, and High Error Rate	4-20
4-8 Regressions Analyzing the Effects of Engaged, Low Error Rate Time, Analyses of Reading and Mathematics Achievement at Grades 2 and 5 during the A-B Period, Regressing Post Achievement on Pre-Achievement and Engaged, Low Error Rate Time	4-22
4-9 Regressions Analyzing the Effects of Engaged Time, Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the B-C Period, Regressing Post Achievement on Pre Achievement and Engaged Time	4-24
4-10 Regressions Analyzing the Effects of Allocated Time and Engagement Rate, Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the B-C Period, Regressing Post Achievement on Pre Achievement, Allocated Time, and Engagement Rate	4-26
4-11 Regressions Analyzing the Effects of Low and High Error Rates, Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the B-C Period, Regressing Post Achievement on Pre Achievement, Engaged Time, Low Error Rate, and High Error Rate	4-28

<u>Table</u>	<u>Page</u>
4-12 Regressions Analyzing the Combined Effects of ALT Variables, Analyses of Reading and Mathematics Achievement in Grades 2 and 5 for the A-B Period, Regressing Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate	4-30
4-13 Regressions Analyzing the Combined Effects of ALT Variables, Analyses of Reading and Mathematics Achievement in Grades 2 and 5 for the B-C Period, Regressing Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate	4-31
4-14 Summary of Regressions, Combined Effects of ALT Variables, Regression of Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate (A-B Period)	4-32
4-15 Summary of Regressions, Combined Effects of ALT Variables, Regression of Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate (B-C Period)	4-33
4-16 Conjoint Relationship of Pre-Achievement (at the A Testing) and Low Error Rate (A-C Period) in Terms of Residual Achievement (C Achievement Residualized on A Achievement)	4-46
5-1 Definitions and Examples of Interactive Teaching Behaviors	5-3
5-2 Frequency of Occurrence of Interactive Teaching Behaviors, by Setting	5-5
5-3 Means, Standard Deviations, Reliabilities and Intercorrelations of Interactive Teaching Behaviors for the B-C Period (Grade 2)	5-8
5-4 Means, Standard Deviations, Reliabilities and Intercorrelations of Interactive Teaching Behaviors for the B-C Period (Grade 5)	5-9
5-5 Correlations of Interactive Teaching Behaviors with % Engaged	5-12
5-6 Variability in Use of Instructional Settings and Average Engagement Rate, by Setting	5-16
5-7 Correlations Between Instructional Behavior and Engagement Within and Across Settings	5-19
5-8 Definitions of the General Variables	5-28
5-9 Means, Standard Deviations and Reliabilities of the General Variables in the B-C Period	5-32
5-10 Intercorrelations of General Variables (Grade 2, B-C Period)	5-33
5-11 Intercorrelations of General Variables (Grade 5, B-C Period)	5-34
5-12 Correlations of General Variables with % Engaged	5-36

<u>Table</u>	<u>Page</u>
6-1 Definitions of Diagnosis and Prescription Variables	6-4
6-2 Means, Standard Deviations, and Reliabilities of the Diagnosis Variables	6-8
6-3 Means, Standard Deviations, and Reliabilities of the Prescription Variables	6-9
6-4 Intercorrelations of Diagnosis and Prescription Variables (Grade 2)	6-11
6-5 Intercorrelations of Diagnosis and Prescription Variables (Grade 5)	6-12
6-6 Correlations of Diagnosis and Prescription Variables with Low Error Rate and High Error Rate	6-13
6-7 Correlations of Interactive Teaching Behaviors with Low Error Rate and High Error Rate (in the B-C Period)	6-22
7-1 Post Achievement, Pre Achievement, and Time Variables Used in A-C Analyses of Teaching Processes, ALT, and Achievement (Grade 2)	7-2
7-2 Post Achievement, Pre Achievement, and Time Variables Used in A-C Analyses of Teaching Processes, ALT, and Achievement (Grade 5)	7-3
7-3 Summary of Correlations Between Planning or General Variables and Residual Achievement (Grade 2)	7-5
7-4 Summary of Correlations Between Planning or General Variables and Residual Achievement (Grade 5)	7-6
7-5 Summary of Correlations Between Interactive Teaching Behaviors and Residual Achievement (Grade 2)	7-7
7-6 Summary of Correlations Between Interactive Teaching Behaviors and Residual Achievement (Grade 5)	7-8
7-7 Correlations of Selected Teaching Process Variables with Facets of ALT	7-13
7-8 Conjoint Relationship of Academic Feedback and Student Engagement to Residual Achievement in Grade 2 Reading	7-21
7-9 Summary of Significant Correlations with Achievement, by Subtest	7-26
7-10 Summary of Regressions Using Teaching Process Variables and ALT Variables to Predict Achievement (Grade 2)	7-29
7-11 Summary of Regressions Using Teaching Process Variables and ALT Variables to Predict Achievement (Grade 5)	7-30

<u>Table</u>	<u>Page</u>
8-1 Number of Second Grade Students and Class Mean Residualized Achievement and Attitude Scores in Reading and Mathematics for 25 Second Grade Classes	8-6
8-2 Table of Intercorrelations of Class Mean Residualized Achievement and Attitude Scores in Reading and Mathematics for 25 Second Grade Classes	8-8
8-3 Class Means on Total Reading and Mathematics for the "A", "B", and "C" Periods for 8 Selected Second Grade Classes	8-10
8-4 Raw Change Scores (Reading and Mathematics) over the "A" to "C" Period for Selected Second Grade Classes	8-11
8-5 General Classroom Variables for 8 Selected 2nd Grade Classes	8-13 8-14
8-6 Time and Instructional Process Variables (Reading) for 8 Selected 2nd Grade Classes	8-16
8-7 Time and Instructional Process Variables (Mathematics) for 8 Selected 2nd Grade Classrooms	8-18
9-1 Scale Range, Standard Deviations, and Alpha Coefficients for Student Attitude Toward Reading, Mathematics, and School at the A, B, and C Testing Occasions	9-4
9-2 Correlations Among Measures of Student Attitude Toward Reading, Mathematics, and School, A-C Period	9-7
9-3 Correlations of Residual Achievement with Residual Attitude, A-C Period	9-8
9-4 Correlations of Summer Retention (Residual Achievement, C-D Testing) with Residual Attitude (A-C Period)	9-10
9-5 Correlations of ALT Variables with Residual Attitude, A-C Period	9-11
9-6 Correlations of Planning or General Variables with Residual Attitude, A-C Period	9-14
9-7 Correlations of Interactive Teaching Behaviors with Residual Attitude, A-C Period	9-15
10-1 Comparison of Achievement Scores at C Between Students Who Were Tested at D and Students Who Were Not Tested at D	10-3
10-2 Mean Achievement on Four Testing Occasions, Grade 2	10-6
10-3 Mean Achievement on Four Testing Occasions, Grade 5	10-7
10-4 Achievement Scores at D Predicted by Three Different Sets of Independent Variables	10-10

<u>Table</u>		<u>Page</u>
10-5	Achievement Scores at D Regressed on Achievement at C, Summer Instruction, Academic Learning Time Variables (AC), and Teaching Process Variables (AC) for Grade 2. (N=97)	10-13
10-6	Achievement Scores at D Regressed on Achievement at C, Summer Instruction, Academic Learning Time Variables (AC), and Teaching Process Variables (AC) for Grade 5. (N=89)	10-15
10-7	Correlations of Academic Learning Time Variables (AC) and Teaching Process Variables (AC) with Residual Achievement at D	10-17

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2-1	The General Model of Instruction.	2-3
2-2	Schematic Configuration of Instructional Functions in the Academic Learning Time Model	2-9
5-1	Plot of Engagement Rate and Task Engagement Feedback Across Weeks for a Single Student	5-25
5-2	Class Engagement Levels Plotted According to Perceived Academic Competence and Orientation Toward Affect (Grade 5, B-C Period)	5-39

CHAPTER 1: INTRODUCTION

The identification of effective teaching behaviors has challenged the educational community for several decades. The task is to conceptualize and assess those instructional factors which promote student learning. Since the context of school learning and the interactional patterns which occur in classrooms are extremely complex, the identification of effective teaching behaviors is not an easy task. Because of this complexity, it is unlikely that a unique set of teaching behaviors will be found to be effective in all situations with a wide variety of learners. Any progress in this area has great potential for the improvement of teacher preparation procedures and school instruction. It is especially important to gain insight into the educational process at the elementary school level, since it is here that learning patterns are established and students acquire basic skills such as reading, writing, and computing. The more knowledge we have about the conditions which foster the acquisition of these skills, the greater our potential for providing appropriate pre-service and in-service training experiences for teachers.

In California, certification of teacher training programs and licensing of teachers are the responsibility of the California Commission for Teacher Preparation and Licensing (CCTPL). Since its inception, the CCTPL has encouraged and supported educational research on teaching effects. The results of research were seen as potentially important contributions to policy in the area of teacher education. In 1972, the CCTPL, with funding from the National Institute of Education, mounted a program of research to identify effective teaching behaviors in the areas of reading and mathematics instruction at grades 2 and 5. The research program, called the Beginning

Teacher Evaluation Study,¹ was conceptualized in several sequential stages or phases. After a general planning year (Phase I, 1972-73), the Commission contracted with the Educational Testing Service to conduct Phase II (1973-74) of the program. During Phase II extensive field work was carried out in California elementary schools to: 1) develop hypotheses about the relations of teaching performances to pupils' learning; 2) develop a measurement system that might be useful in subsequent phases of the BTES project; and 3) estimate the relative influence of teachers' characteristics and the environment of teaching on teaching performance and of pupils' characteristics and background on pupils' learning (McDonald & Elias 1976, p. 1). The results of Phase II (McDonald & Elias, 1976) were to be used as one of the major starting points for subsequent phases of the program.

Phase III of the research program was conducted by the Far West Laboratory for Educational Research and Development. During Phase III (1974-78), an initial period (Phase III-A, 1974-75) was spent developing additional hypotheses. Results of this work, Phase II, and other research on teacher effectiveness were integrated in the form of a model of instruction. Refinement of this model and development of appropriate instrumentation was carried out during the 1975-76 school year (Phase III-A Continuation). The model, referred to as the Academic Learning Time (ALT) model of classroom instruction, was empirically evaluated in an extensive

¹ The title of the research program indicates a focus on beginning teachers. In early planning for the program (Phase I), the beginning teacher concept played a relatively prominent role. For a number of reasons, including rapid reduction in the rate of entry of new teachers into the field, the study was not limited to beginning teachers. The title of the research program, however, remained unchanged throughout all phases of the study. The teachers participating in Phase III of the study varied in teaching experience from first year teachers to veterans of more than twenty years.

correlational field study (Phase III-B,² 1976-78).

This document is the final report of the Phase III-B correlational field study. The progress of Phase III-B is documented in a series of technical notes (Appendix 1-A). The technical notes cover instrumentation, description of field procedures, testing results, generation of descriptive statistics, and analysis of major relationships. This report, which draws heavily on the series of technical notes, is intended to provide a technical summary of the Phase III-B correlational study. Those interested in a more detailed description of procedures and instrumentation may refer to the series of technical notes. A non-technical summary of Phase III (Fisher, Berliner, Filby, Marliave, Cahen, Dishaw and Moore, 1978) is also available.

In many ways, each subsequent phase of the study built upon previous phases culminating with the Phase III-B correlation study. However, the overall progression within the study was not linear. The central variables and the final design for Phase III-B evolved in a complex manner. A sense of the diversity of issues addressed within Phase III³ may be obtained by examining the annotated references in Appendix 1-A.

Context for Phase III-B

The systematic study of teacher effects is a relatively new field. Only in the past few decades has there been concentrated effort to plan and carry out studies specifically to identify the impact of teaching on

²Phase III-B also included an intervention study in which the Academic Learning Time (ALT) model was used to guide specific interventions in a subset of four grade 2 classrooms (Berliner, Filby, Marliave, & Weir, 1978).

³The data collected during Phase III of BTES are available for secondary analysis. Information on documented data tapes may be obtained from the Far West Laboratory for Educational Research and Development.

student learning (Dunkin & Biddle, 1974; Gage, 1963; Travers, 1973). Like any new field, research on teaching effects has been characterized by sharp differences in methodology and generally large amounts of error associated with its measurement systems. Given the complexity of the situation -- one or more adults in a classroom with approximately thirty students and almost continuous activity over long periods of time -- it is not surprising that simple and generalizable findings have not readily been found.

Early studies of teacher effects attempted to relate static characteristics of teachers (number of years experience, number of preservice courses, etc.) to indices of student learning (achievement scores). Gradually studies began to focus on the classroom behaviors of teachers and attempts were made to relate frequency or quality of these teacher behaviors to student achievement. As attempts were made to identify more specific relationships, the description of teaching behavior became more and more molecular. That is, smaller and smaller slices of behavior were described and related to student achievement. As a result, the list of "important" teaching behaviors grew longer and longer. Although relationships were found, they appeared to be unstable. Relationships determined in one setting were, in general, very difficult to replicate even with small changes in context. The identification of stable relationships was further complicated by the probability of interactions among instructional factors and student characteristics.

Most studies of teaching used student cognitive achievement as the major criterion for student learning. Student achievement was invariably assessed with paper and pencil tests. Partly because of the apparent instability in relationships between teaching behaviors and student test scores (but for many other reasons as well), achievement tests came under

heavy scrutiny. In some cases, items included on a test did not match the content covered during instruction. In addition, the feedback value of achievement test scores, for the purpose of identifying effective teaching behaviors, was questioned. If a test is readministered after a one-month interval, it provides a measure of what a given student has learned during that time. However, during that time literally thousands of teaching acts were performed. Presumably some of the teaching acts were more beneficial than others, but the achievement test score sheds no light on this point. Since achievement test scores do not reflect behaviors which are co-extensive in time with classroom instructional acts, the scores provide little or no information about differential effectiveness of these acts.

Inconclusiveness in earlier studies of teaching effectiveness and awareness of difficulties with traditional criterion measures of student learning, helped turn attention to the student and his or her behavior during instruction. Studies of teaching gradually became more complex. The chain of influence on student learning was traced from static teacher characteristics to teacher classroom behavior to student classroom behavior to indices of student learning.

It was in this general context that Phase III⁴ of the Beginning Teacher Evaluation Study was carried out. During Phase III-A, experimental teaching units were developed and implemented in approximately one hundred classrooms at each of two grade levels. Each teacher taught a two-week unit in reading and one in mathematics. These units allowed the content (objectives) of instruction to be controlled while giving the teachers considerable freedom in how they designed and implemented their instruction.

⁴The technical reports describing the work completed as part of Phase III are listed in Appendix 1-A.

Analysis of student pre and posttest scores, given in conjunction with the experimental teaching units, generated a distribution of more and less effective teachers. Teachers from the extremes of the distribution were subsequently observed and interviewed in a series of special studies. One study examined the regular classroom teaching of the more and less effective teachers using ethnographic methods. Additional special studies examined the instructional planning behavior of more and less effective teachers and student perceptions of the teachers. During Phase III methodological work on the assessment of stability of interactional behaviors was also completed.

While this field work was being conducted, a number of review and instructional model-building activities were completed. A key element in the modeling activity focused on student learning time and on the assessment of instructional processes in terms of elapsed time. This work drew heavily on Carroll's model of school learning (Carroll, 1963) and subsequent work on time and learning by Bloom (1964, 1973), Harnischfeger and Wiley (1975), and Wiley (1973).

During Phase III-A Continuation we used the results of previous phases as well as other research on teaching to address some of the major issues in research on instructional effectiveness. Gradually we developed a crude model of classroom learning. Field work was carried out to develop and refine instruments for collecting classroom process information in terms of time. Achievement tests were further developed and tested in the field in order to identify items and scales which were reactive to instruction. Time in instruction was coded in a series of mutually exclusive content categories which were matched to the achievement subtests. In addition, instruments were developed for the direct observation of student engagement. A sub-study examined the relationship between engaged time and student achievement.

Phase III-B

On completion of this work, the concept of "Academic Learning Time" (ALT) became a major focus of our work. Academic learning time is intended to be a quantitative measure of ongoing student learning based on observable classroom behavior. It is defined as the amount of time a student spends engaged on a task that produces few student errors and which is directly related to a defined content area. The central hypothesis of the ALT model of classroom instruction is that one can measure ongoing student learning in the ALT metric and that this ongoing measure of learning provides a new and better way to identify and understand effective instructional processes. More specifically, we hypothesized that the amount of ALT a student accumulated would be related to the more traditional measure of student learning -- achievement test scores. Furthermore, we hypothesized that instructional process variables, as described in the next chapter, would be related to ALT. That is, teaching processes should influence student learning behavior in the classroom. The different components of ALT, such as student engagement and student error rate, may allow us to pinpoint instructional effects.

Phase III-B was designed to evaluate this instructional model. An extensive longitudinal process-product study provided the data base for the evaluation. Student achievement was assessed at three points during a school year. Large amounts of instructional process data were collected by direct observation, teacher logs, interviews, and ratings during the inter-test periods. This information was used to address the first two major questions of the study:

1. What is the relationship between student academic learning time and student achievement?
2. What are the relationships between teaching processes and academic learning time?

Measures of student attitude toward reading, mathematics and school were administered at the three major testing points. These data were analyzed in conjunction with instructional process and student achievement measures. In addition, a fourth achievement testing was carried out at the beginning of the following school year. Analyses of these two sources of information addressed the following general questions:

3. Are teaching processes, academic learning time, and student achievement related to student attitudes?
4. Are instructional variables related to retention of achievement over the summer?

Structure of the Report

The remainder of this report describes the data collection and presents analyses and results relevant to the four general questions above. Chapter 2 presents the Academic Learning Time model of classroom instruction. The design, instrumentation, and data collection activities are described in Chapter 3. Chapter 4 presents analyses of academic learning time and student achievement. Chapters 5 and 6 describe relationships between teaching processes and academic learning time (Chapter 5 focuses on student engagement, Chapter 6 focuses on student error rate). Chapter 7 contains an analysis of the joint relationships among the three major sets of variables: teaching processes, academic learning time, and student achievement.

The analyses in Chapters 4 through 7 focus on quantitative relationships among variables. In Chapter 8, specific classrooms are analyzed in a more qualitative fashion. Data collected from a wide variety of sources provide concrete descriptions of classes in which students showed relatively high gain in achievement and classes which showed relatively low gain in achievement. Chapter 9 presents analyses of student attitude while Chapter 10 describes analyses of retention of achievement. A summary of results and conclusions of the study are contained in Chapter 11.

CHAPTER 2: THE ACADEMIC LEARNING TIME MODEL OF CLASSROOM INSTRUCTION

Introduction

During the course of Phase III of BTES, a general framework for conceptualizing classroom instruction and student learning evolved. This framework or model has been revised in a number of ways as the study progressed. The purpose of this chapter is to describe the central concepts of the model and to discuss the relationships among the concepts. Information is provided on the recent history and evolution of the model. This material leads to a description of the ALT model as we currently view it; however, further specifications and elaboration of the model are likely to occur as the result of practical experience and additional research. In later chapters, a number of implications for future development of the model are discussed.

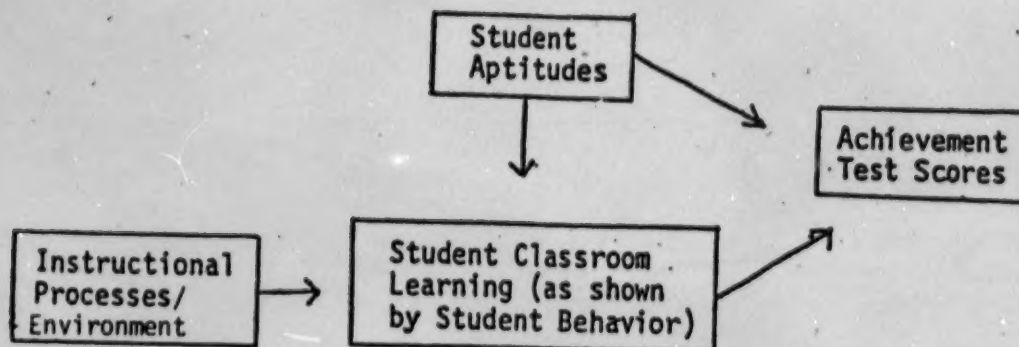
The general goal of classroom research on teaching effectiveness is to identify classroom conditions and teaching acts which foster student learning. What is it that teachers do that promotes student learning? In many empirical investigations of this question, achievement tests have been used as the measure of student learning. Although performance on achievement tests may be important for many reasons, test scores have a number of disadvantages when used in quantifying teaching effectiveness. For example, it is very difficult to separate the effects of out-of-school experiences on test scores from the effects of school instruction. In pre-post testing situations, achievement tests are usually readministered after several days or weeks of instruction. During the intertest period, many teaching acts are performed. As a result, it is impractical to use feedback from test scores to differentiate the effectiveness of one teaching act or sequence from another.

Achievement test scores provide summary information for the entire instructional period. In fact, achievement test scores tend to ignore the learning process itself. Student learning takes place in individual student minds as they work in classrooms throughout the school year. Scores on achievement tests provide one index of student learning, but they do not constitute learning itself.

To precisely quantify or "capture" learning is impossible, but it may be quite possible to assess more proximal indices of learning than those provided by test scores. If a more proximal index of student classroom learning could be obtained, then it should provide a useful criterion for the study of teaching.

Student behaviors during instruction should provide a more immediate indicator of learning as it takes place. Instructional process variables, then, influence ongoing classroom learning behavior. Over time, a summary index of this learning is provided by test scores. The crucial point is that student classroom learning behaviors and subsequent test scores both constitute indices of learning. However, the classroom learning behaviors are more proximal than achievement test scores to the instructional process. Figure 2-1 presents a schematic description of these concepts and their interrelationships. The general model of instruction states that, for a given student, instructional processes (what happens in the classroom) lead to student learning behaviors which subsequently lead to achievement test scores. Student aptitudes also have a direct impact on both student learning behaviors and achievement test scores. These aptitudes represent characteristics of the student which are "brought to" the learning situation.

Figure 2-1
The General Model of Instruction



At this stage, the instructional process/environment portion of Figure 2-1 includes all teaching behaviors, instructional activities and contextual conditions in the classroom. We will elaborate on our conception of teaching processes and context variables in a later section of this chapter. First, examine the central element in Figure 2-1, student classroom learning behavior.

Student classroom learning was operationalized in terms of the amount of time students spent working on academic material. The definition of learning time changed over the course of the study. We began by examining the amount of time available during school for a student to work in a particular subject area. In grade 2 reading, it is common to teach decoding. For each pupil, a certain portion of the school day is set aside for work on decoding. This kind of structuring of time is usually, but not always, done by the teacher. In any case, this block (or blocks) of time constitute the maximum time available for the student in question to work on decoding. We refer to this

quantity as time allocated to decoding. However, all of the time allocated to decoding will not necessarily be used for work on decoding. For some of that time the student will be actively engaged in instructional activities relevant to decoding but, for some of the time, he will be off-task or unengaged. Hence engaged time represents a somewhat more refined measure of student classroom learning than allocated time.

The difficulty of the decoding task for the student will also influence the amount learned. If the task is very difficult and the student produces many errors among his responses to the task, then the activity will not produce much learning. On the other hand, if the student produces few errors on the task, then we hypothesize that learning is occurring. The student has to make a correct response (read the word correctly or get the right sum) before s/he can learn and remember the correct response.

Time spent by a student engaged on a task on which few errors are produced, and where the task is directly relevant to an academic outcome, constitutes a measure of student classroom learning behavior. We refer to time spent under these conditions as Academic Learning Time (ALT). Thus, the basic components of ALT are allocated time, student engagement, student error rate, and task relevance to an academic outcome. The ALT model hypothesizes that the accumulation of academic learning time will lead to gains in achievement.

Although ALT is a unitary concept, we have defined it in terms of several component elements (allocated time, engagement, error rate, and relevance to a content domain). It is convenient to think about ALT as one entity or variable; however, in the conduct of the study we measured and analyzed the components separately. In later chapters, results are presented for each of the components. In some chapters, the results for the

components are summed and presented as results for ALT. The following paragraphs provide some elaboration on each of the facets of ALT.

Engagement refers to the involvement of the student in the learning activity under consideration. Obviously, if the student is not paying attention, no learning can occur. Therefore, we distinguish between being engaged (apparently paying attention) and not being engaged. This distinction between student engaged time and teacher allocation of time is critical. Allocated time is a function of teacher behavior, the assignment of a given activity to a specified period of student time. This is important, in that, it generally sets the limit for student engaged time in that particular activity. Nevertheless, there may be vast differences between allocated time and student engaged time. These differences may occur generally for an entire class or differentially across students within a class. Because student engagement in a learning activity is a necessary precondition for learning, ALT must include only engaged time.

Error rate, as used in the ALT model, is a complex concept. It is intended to reflect the degree to which the student correctly processes and understands the learning task. When a student works on a specific task, there is an interaction between the characteristics of the task and the current ability of the student to complete the task. A given student will learn more from some tasks than from others. Hence there are degrees of appropriateness among tasks which affect student learning. For a given student, periods of time spent on tasks differing in this kind of appropriateness will be quite different in quality and will affect the amount the student learns from doing the task. Error rate is intended to reflect this quality in the periods of time a student spends working on school tasks.

Note that the error rate concept is an interactive one; that is, it depends upon both the student and the instructional task. In earlier work on the BTES, we referred to this interaction between task and student as task difficulty level. For a given student, we assessed whether a particular task was easy, medium, or difficult for that student. However, the term task difficulty level sounds like a characteristic of the task only. Because of this somewhat misleading connotation, in later work we use the term student error rate. The latter term places more emphasis on the behavior of the student and is therefore preferred, but it does not entirely convey the interactive nature of the concept.

In this study, error rate is assessed in three fairly broad categories. Low error rate describes situations where the student has a good grasp of the task and makes errors at about the chance level ("careless" errors). If a student does not understand the task and makes correct responses at about the chance level, then this situation is labelled high error rate. Situations which fall between low and high error rate are defined as medium error rate. Medium error rate involves a situation of partial knowledge, where the student understands enough to produce some correct responses, but also commits errors due to limitations in his/her understanding of the task. In an earlier version of the ALT model, we hypothesized that medium error rate would be most likely to reflect student learning and that high error rate would be detrimental. Based on preliminary data analysis, the hypothesis held for high error rate; however, low error rate was positively related to achievement gain while medium error rate was not. Apparently, tasks which produce low error rates provide situations where students can rapidly improve their performance and continue to learn as tasks with small increases in difficulty are encountered. Hence, the current ALT model states

that more time spent on tasks which produce a low error rate leads to increased achievement. However, this does not necessarily imply that all of a student's time should be spent in the low error rate condition. Part of the original belief in medium error rate was based on the idea that some mental effort is required to produce learning. If a task is so easy for a child that s/he does it without thinking about it, then little new learning is likely to occur. Generally, it is expected that some balance between a low and a moderate error rate, with somewhat more activities at a low error rate than at a moderate error rate, will produce the most student learning. A high error rate would always be detrimental to learning conditions.

Task relevance to an academic outcome, as defined in the current ALT model, simply identifies the content category to which ALT is directed. A particular study might focus on a narrow range of content outcome categories, and therefore a narrow range of ALT. Clearly, if one attempts to relate ALT to gain on an achievement test battery, then ALT must be limited to those content categories that are relevant to that battery. The present study is limited to ALT involving tasks that are directly relevant to reading and mathematics outcomes. We are not assessing the effects of ALT for other academic outcomes. Conceptually, however, ALT could be applied to any academic outcome.

The absence of ALT is also important. A great deal of student time in class consists of "Non-ALT" because the student is not engaged in any academic task or because the academic task in which the student is engaged produces an inappropriate error rate. The usefulness of the concept, "Non-ALT," lies in the specification of its cause, either lack of engagement or an unreasonable student error rate. Hence, it is possible to analyze, in terms of student variables, why learning does not occur.

A major strength of the ALT concept is that it focuses on an individual student and his ongoing learning behaviors. The model is remarkably flexible, in that, a teacher may keep track of engagement and error rate for groups or for the whole class by "averaging" over individual students. However, the major advantage of the model is that it helps a teacher really "see" what individual students are doing in terms of their observable classroom learning behaviors. The power of this perspective lies in the fact that learning is really an individual phenomenon, not a group phenomenon. Therefore, while managing the class as a group, the teacher must have some perspective from which the ongoing learning of individual students can be monitored.

Chapter 3 provides information on the methodology used to assess the components of ALT. Further discussion of ALT and results for the analysis of ALT and student achievement are presented in Chapter 4.

Instructional Processes

ALT represents a proximal measure of student-learning-as-it-occurs. The instructional processes that result in increased student ALT, will also yield increased student learning. Therefore ALT is central to the instructional process, and serves as the basis of the present analysis of teaching behaviors.

The teaching behaviors that influence ALT can be conceptualized in terms of five interrelated functions (diagnosis, prescription, presentation, monitoring, and feedback) that occur through time in a roughly cyclical fashion, as shown in Figure 2-2.

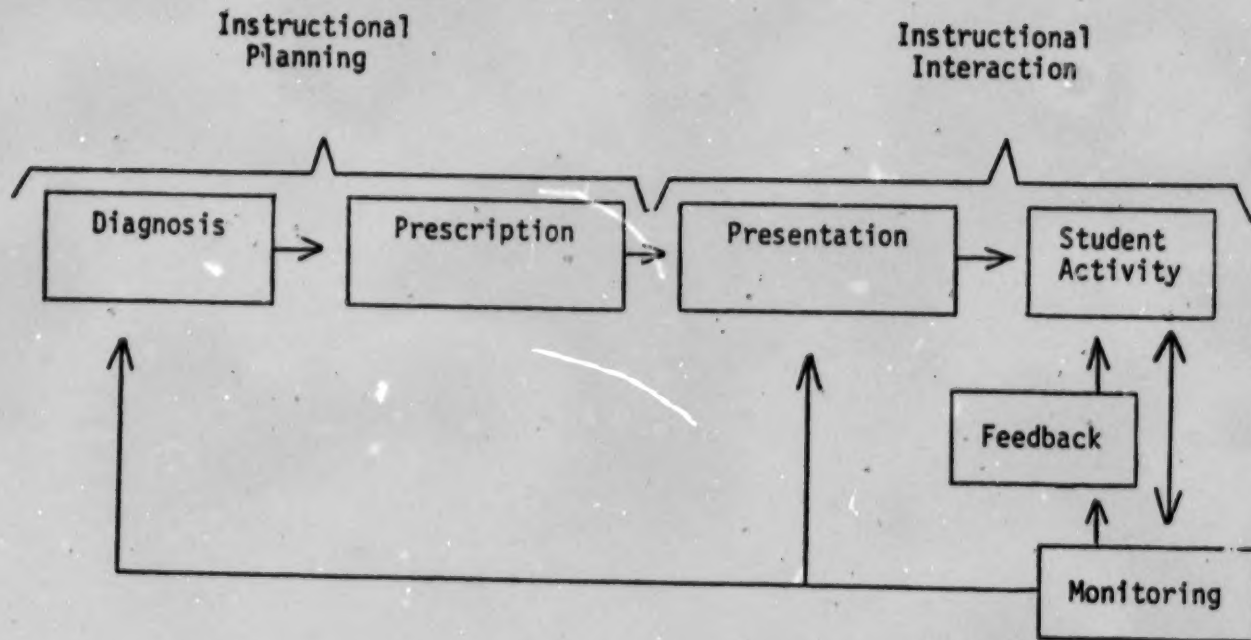


Figure 2-2

Schematic Configuration of Instructional Functions in the Academic Learning Time Model

Depending upon the class organization and curriculum characteristics, the functions in the basic cycle may be satisfied by apparently very different behaviors. Monitoring in a large group discussion is very different from monitoring during seatwork. In the first case, monitoring is implicit in the process we call discussion. In the second case, monitoring may or may not be performed by the teacher at any given time. Similarly, individual teachers may vary in the behaviors with which they fulfill these functions.

The cycle of functions begins with a planning phase. The teacher as organizer and decision-maker assesses the current knowledge, skill levels, strengths, and weaknesses, of the students (diagnosis). S/he can then decide

on appropriate instructional activities and methods for reaching those goals (prescription). These decisions set the stage for the instructional interaction phase.

The interaction phase begins with the presentation of concepts or learning tasks to the student. The student responds to this input in some way, with an overt action, with a covert assimilation of information, or by ignoring the input. The teacher monitors the student's responses in order to know whether the instructional goal is being achieved. Monitoring tells the teacher about the student's state of knowledge or skill following an instructional presentation. Guided by information from monitoring, the teacher may provide feedback to the student, provide additional explanation or cycle back to the beginning for further diagnosis and planning.

The focus on the teacher as a conscious decision-maker occurs throughout this cycle. The teacher arrives at goals, presents instruction toward those goals, and evaluates whether or not the goals are being reached. S/he decides whether a student's response is adequate or inadequate, and decides, on the basis of the student's response, what to do next.

ALT provides the vehicle for monitoring the impact of this instructional process. Each of the functions in the model contributes directly to establishing or maintaining ALT. Accurate diagnosis and prescription enable the teacher to assign tasks of an appropriate difficulty level for the student, in a content area that is relevant to the student's needs.

Presentation that is well structured, clear, and relevant to student needs results in greater student understanding of the task. This, in turn, produces higher engagement and an appropriate student error rate. Monitoring and feedback maintain student engagement and understanding of the task, and allow the teacher to rediagnose and represcribe when appropriate. An examination

of ongoing instruction in terms of student ALT and related teaching functions allows one to identify when and where the student learning process is being interrupted, relating these interruptions to instructional functions.

The methodology used in assessing the teaching behaviors is described in Chapter 3. Specific variables are described and results presented for interactive behaviors and planning behaviors in Chapters 5 and 6 respectively.

Classroom Environment

Classroom environment plays an important role in instruction. The functional model of teaching processes depicted in Figure 2-2 operates within classroom environments. Classes may differ widely on such dimensions as enthusiasm, warmth, competitiveness, cooperation, and task orientation. Differences in environmental variables or classroom context variables may influence academic learning time directly. For example, the value system of the teacher constitutes part of the classroom environment. Some teachers may value academic pursuits very highly and hence provide a classroom environment which has high academic press. This press may tend to raise the general level of engagement in the classroom.

Classroom environment may influence academic learning time via a second mechanism. Differences in environmental variables may mediate the relationship between teaching process variables and facets of academic learning time. For example, the specific behaviors which constitute feedback might be very different in classrooms where the climate differed in warmth. In a high warmth class, the information on correctness of student response might be accompanied with a smile or a reassuring touch, while in a less warm class the feedback might be given without expression of affect. In both cases, the feedback function would be fulfilled; however, it is possible that the difference in warmth would moderate the relationship between feedback and

student engagement.

Generally, the cyclical five function model should hold in many different classroom environments, however, differences in classroom environments may have direct effects on academic learning time. In some cases, environmental variables may moderate the relationship between teaching processes and academic learning time. In this study, classroom environment has been assessed from ratings and interviews. Specific variables are described and results presented in Chapter 5.

The ALT model or conceptual framework has a number of important characteristics. The central role in the model is given to academic learning time. By doing so, the model begins to "unpack" the complex relationships between teaching behaviors and student test scores. Academic learning time provides an index of ongoing student learning as it occurs.

Another advantage of the ALT concept lies in the fact that it is coextensive in time with instruction itself. Hence, it provides a proximal student variable against which assessments of ongoing classroom conditions and behaviors can be made. The feedback value to a teacher, provided by ALT, is potentially very great. An awareness of the ALT for an individual student or the profile of ALT across the students in a class may not only help a teacher to decide when to intervene on an instructional sequence but also provide strong indicators of what to change. This is an important point, since this framework provides an observable in-class criterion which can guide minute to minute instructional decisions.

The model also conceptualizes teaching behaviors in terms of the functions they serve in the instructional setting. When teaching behaviors are defined at a very molecular level, it is extremely difficult to show stability in the impact of the behavior over even relatively small changes in

context. It appears that the same behavior can serve different functions depending upon the context. This implies, for example, that there is no one specific behavior which will turn out to be the essential performance of the monitoring function. This functional view of teaching behavior has high potential for conceptual elaboration and especially for the practice of teaching. Certainly teachers need a repertoire of specific teaching behaviors, but they must also have a good grasp of the functions which specific behaviors fulfill.

Summary

The academic learning time model of classroom instruction proposes relationships among three sets of variables: teaching process variables, student classroom learning behaviors (ALT), and student achievement. All variables are conceptualized at the level of individual students. The teaching process variables are conceptualized in terms of five interrelated teaching functions (diagnosis, prescription, presentation, monitoring, and feedback). Academic learning time constitutes the classroom learning behaviors. ALT is operationalized as time spent engaged on a task on which the student produces few errors. The task must be directly relevant to a desired academic outcome. Achievement refers to learning as assessed by paper and pencil achievement tests. The model states that, for a given student, the accumulation of ALT will result in achievement gain. Secondly, if the five teaching functions are fulfilled for the student during his or her ongoing instruction, then the accumulation of ALT will be enhanced. The model also hypothesizes that student entering characteristics (aptitudes) affect both academic learning time and achievement. Teaching processes are viewed as occurring within specific classroom environments. Characteristics of the environment may relate directly to ALT or may moderate the impact of teaching process variables on academic learning time.

CHAPTER 3: OVERVIEW OF DESIGN AND INSTRUMENTATION

Design of the Correlational Field Study

The design of the Phase III-B field study called for student achievement measures to be collected at three points in time with extensive collection of instructional process information during the inter-test periods. The achievement measures covered specific content areas in reading and mathematics. Data were collected during the 1976-1977 school year (with follow-up achievement testing in September, 1977) on two samples of students -- one at grade 2 and one at grade 5.

Instructional process data were collected by a trained field staff during the two major inter-test periods. The first or A-B period began with the A test administration in October, 1976 and ended with the B test administration in early December, 1976. There were six weeks between the A and B test points. The second or B-C period began with the resumption of school in January, 1977 and ended with the C test administration in April, 1977. The B-C period spanned 17 weeks and, together with the A-B period, provided 23 weeks of instructional process data. Extensive process data were not collected after the C testing occasion. (Follow-up testing, occasion D, was conducted in September 1977).

During the A-B and B-C inter-test periods, two sets of instructional time data were collected. One set contained information submitted by teachers for every school day. These data, referred to as teacher logs, provided daily estimates of the amount of time allocated to individual students within specific content categories. The second set of data on instructional time was obtained by direct observation. Trained field workers observed students for one day of each week during the inter-test periods. This procedure

provided information for individual students on time allocated to specific content categories, student engagement rates, and error rates for each student.

The observation procedure also provided information on teaching processes. Information was gathered for individual students on interactive teaching processes such as presentation, monitoring, and feedback. Interviews, ratings, and self report measures provided information on diagnosis and prescription as well as teacher aptitude and classroom environment.

Sample

Teachers were recruited during the spring of 1976. All teachers participating in the study were volunteers. The recruitment was carried out by the California Commission for Teacher Preparation and Licensing in several school districts in the San Francisco Bay Area. This area was selected because it included a variety of urban and suburban settings and was easily accessible from the Far West Laboratory. School districts were contacted and given information about the study. With district approval, building principals and teachers were contacted. At school meetings the overall objectives of the study and obligations of participants were presented. A pool of approximately 50 teachers at each grade level was obtained. Teachers who participated in the full schedule of study activities were given small honoraria and were eligible for extension course credits through a cooperating teacher training institution. A complete description of recruiting procedures is presented in Technical Note II-1, Phase III-B (Howell & Rice, 1977).

In order to minimize ceiling and floor effects on repeated achievement testing, and to allow collection of process data on individual students our design prescribed the selection of six target students within each class.

This selection was carried out by administering a representative battery of reading and mathematics subscales to each student in the volunteer classes. Forty second grade classes and 41 fifth grade classes participated in the testing (September, 1976). From the distribution of scores for each grade we established a middle range band from which to identify target students. Target students came from a low middle range on the overall distribution of achievement scores (the distribution for all classes combined). Almost all target students fell between the 30th and 60th percentiles on both reading and mathematics distributions. On the basis of this achievement information and teacher feedback, we identified three boys and three girls in each class. Twenty-eight second grade and 30 fifth grade classes met our criterion of having six or more students falling within our selection band of scores in reading and mathematics. The teachers of these classes were invited to be study participants. A description of the selection testing and target student selection procedures is presented in Technical Note III-1 (Cahen, 1977). Prior to the A testing sessions some teachers (three at second grade and eight at fifth grade) reconsidered their commitment and withdrew from the study. Almost all of the teachers who withdrew did so because they were unable to make the time commitment necessary for participation.

At the beginning of the A-B period there were 25 second grade and 22 fifth grade classes participating in the study. All but one of these classes remained in the study throughout the A-B and B-C periods. (One fifth grade class was dropped during the A-B period due to non-submission of log data).

The sample of teachers were all volunteers. However, there is little reason to believe that the sample is exceptional. The group contained more women than men (see Appendix 3-A), but that is expected in the population of elementary school teachers. The sample was ethnically mixed; varied

considerably in age and years of teaching experience; and in our experience, covered a considerable range of teaching style and ability.

It should also be pointed out that the teachers were not aware of the ALT model. They were not instructed on the relationships hypothesized by the model nor on the conceptions of teaching behavior used in the study.¹ The teachers were aware that the study was examining teaching effectiveness. They also knew that time allocation was germane to the study, since they were trained by Far West Laboratory staff on log keeping procedures. They also knew who the target students were since they kept logs for only those students. We have no reason to believe that the behavior of these teachers changed significantly because of their participation in the study. The debriefing notes of the field workers were examined for comments which might bear on this point. There were only a few reports of teachers "trying harder" or singling out target students for special attention, and these were minor.

It should be noted that even if there were deliberate changes in teacher behavior due to participation in the study, it would not endanger the results. If teachers allocated more time to target students, that would be noted in the teacher logs and by the direct observation system. Similarly, if teachers initiated more interaction with a target student, the behaviors would be recorded by the observation system. Since our interest is in examining the relations between variations in teaching behaviors and indices of student learning, such interventions by the teacher would be taken into account as part of the measured behavior in that classroom. All things considered, the sample of teachers appeared to represent a cross section of elementary school teachers. Furthermore, their possible knowledge about the ALT model, in all

¹During the B-C period several grade 2 teachers participated in an intervention study (Technical Report VI-1). The ALT model was discussed explicitly with these teachers.

probability, did not alter the results.

The classes of these teachers provided an initial student sample of 150 second and 134 fifth grade students. By the end of the B-C period the student sample had dropped to 139 second graders and 122 fifth graders. The reduction in the student sample was due, in the main, to student mobility. For analysis purposes students who were lost during the B-C period, but present during the A-B period, were deleted from A-B analysis. Thus our final sample included only those students who completed the A-C period (Grade 2, N = 139; Grade 5, N = 122).

The student sample contained approximately 54% girls at grade 2 and 43% girls at grade 5. Non-white students constituted approximately 40% and 30% of the samples at grades 2 and 5 respectively (see Appendix 3-A). Crude information on student socio-economic status was collected by requesting teachers to categorize parent occupations into one of four categories (see Appendix 3-B). This information was collected for both target and non-target students. When all students were considered, approximately 10% fell into group A (executive, professionals, managers), 28% fell into group B (semi-professional, clerical, sales workers, technicians), 47% fell into group C (skilled and semi-skilled employees) and 15% fell into group D (unskilled employees). The sample of target students was distributed very similarly. That is, the target students did not misrepresent the overall distribution of their classmates on socio-economic status.

The target students at both grade levels represent mainstream students. The sample was evenly divided on numbers of boys and girls; the students were neither very high nor very low on entering achievement; the sample on the whole was ethnically mixed; and the distribution of the sample on socio-economic status was very similar to that of the population of students from

which they were drawn. On the whole there is no reason to believe that this group of students is unusual. It is very likely that a similar group of students could be found in and around practically any large North American urban area.

Student Achievement Measures

Student achievement was assessed with the BTES reading and mathematics test battery. The battery was administered in October, 1976 (occasion A), December, 1976 (occasion B), and May, 1977 (occasion C). Each administration required approximately 180 minutes of testing time and was carried out in four 45 minute sessions. Two sessions were completed on each of two days of testing. At least one non-testing day occurred between the two testing days at each administration of the battery. The majority of the items were in the multiple choice format. Information on the development of the battery, reactivity analyses of the items and subtests, and the items themselves are included in BTES Technical Report III-6, Phase III-A Continuation (Filby & Dishaw, 1976).

Student achievement was also assessed in September, 1977 (occasion D). This follow-up testing used a subset (approximately half) of the scales from the BTES reading and mathematics test battery.

Item responses were corrected for guessing (Thorndike, 1971). With few exceptions, subscales administered at A were repeated at B and C. Past experience indicated that instruction relevant to some scales occurs early in the school year while instruction relevant to other scales occurs later in the school year. We attempted to administer subtests which were appropriate for each period. The names of the subscales for each of the 4 quadrants (grade 2 reading, grade 2 mathematics, grade 5 reading, and grade 5 mathematics), the number of items, and reliability coefficients for each subscale

are included in Appendix 3-C. This appendix also includes means and standard deviations for each scale at each of the testing points.

Additional psychometric work on the BTES test battery was undertaken as part of a separate contract to Educational Evaluation and Research Incorporated. Estimates of subscale reliabilities and errors of measurement (Wright & Kim, 1977) were obtained from an analysis of field data collected during April and May, 1977. The BTES battery was administered twice in a two-week period to a sample of students from classes similar to those included in Phase III-B of the BTES. Reliability information on the subscales, taken from Wright and Kim (1977), is included in Appendix 3-C.

In preparation for analysis, data on the subtests were carefully examined. Several subtests were omitted from further consideration based on factors such as low gain, low time allocation in the matched content category, low internal consistency, low test-retest reliability, and ceiling or floor effects.

In the analysis of time and achievement we examined content areas of varying breadth. For this reason we created several subtest combinations. The subtests and subtest combinations used in the analysis of time and achievement are also listed in Appendix 3-C. Of the 13 subtests in grade 2 reading, 11 were used to generate 6 scales for analysis (see list in right-most column of Table 3C-21, Appendix 3-C). Eleven of 12 grade 2 mathematics subtests were retained and aggregated into eight scales for analysis. At the fifth grade level, 9 of 11 reading subtests generated 5 scales for analysis; and 12 of the 14 mathematics subtests were used to generate seven scales for analysis. These "scales for analysis" were used as dependent variables in the analysis of academic learning time and student achievement. In the majority of cases, pre and posttest scores on these variables were available. In a few cases, a pretest score on an "analysis" scale was not available because some

47

subscales that were administered at B were not administered at A. In these instances, A subscales which were logically related to the "analysis" scales were used as pretests.

The total reading (and mathematics) scale was made up of all the items given at a particular testing occasion. Hence the items forming total reading at B for grade 2 are not identical to the items constituting total reading at A or C for grade 2. However the overlap in items forming these total scales is very large.

Student Attitude Variables

Attitude items grouped into three scales (attitude toward reading, attitude toward mathematics, and attitude toward school) were administered at the A, B, and C testing occasions. At grade 2, students responded to each item on a three point scale. At grade 5, each item response was made on a seven point scale. In all cases, item scores were summed to yield scale scores. Descriptive information for the attitude scales for the A, B, and C occasions is presented in Appendix 3-C (Tables 3C-23 through 3C-25).

Academic Learning Time Variables

Measures of the components of academic learning time (allocated time, engagement rates, and error rates) were obtained from two sources. Teacher logs provided a measure of the time allocated to specific content categories. Direct observation provided data on engagement and error rates as well as a measure of allocated time.

Allocated time from teacher logs. Teachers recorded the time allocated to reading and mathematics instruction for each target student for each day in the A-B and B-C periods. Times in minutes were recorded in specific content categories. A list of the log content categories for each grade is

included in Appendix 3-D. A description of the data collection procedures, initial data reduction activities, and descriptive statistics associated with the teacher logs are contained in Technical Note IV-2a (Dishaw, 1977). Daily entries of allocated time were summed for all days in the A-B and B-C periods yielding total time in minutes for each specific content category for each student. These sums were prorated to account for the few cases where daily teacher logs were missing.

Allocated time, engagement rates and error rates from observation. The components of academic learning time were measured by direct observation of target students in participating classes. The observation system used in BTES is a time-based, rotating sample procedure. An event for each target student is sampled about once every four minutes. Each event is coded on several dimensions. One dimension is content, including various subcategories of reading and mathematics. A second dimension is student error rate, where a distinction is made between "low," "medium," and "high." A third dimension is student behavior, including several types of "engaged" and "non-engaged" behaviors. Appendix 3-E gives brief descriptions of the categories in the observation system.

Observation was conducted for a full day approximately once each week. During the A-B period, most classes were observed five times; during the B-C period fifteen times. For analysis purposes, time in each category was summed on each observation day. The observation procedure provided measures of time allocated to each content category for each target student.

Details on instrument development can be found in Technical Report I-5 (Marliave, Fisher, Filby, & Dishaw, 1977). Details on data collection, initial processing, and descriptive results can be found in Technical Note IV-1a (Filby & Marliave, 1977) and Technical Report IV-1b (Fisher,

Filby, & Marliave, 1977).

Two observers were assigned to each class. Observers alternated classes so that each collected data in a given class every other week. Each observer was assigned to eight classes and was paired with different observers in different classes. This meant that each observer had to be familiar with eight classes and about 48 target students. The observation of each class by two different observers reduced the problem of observer bias while still allowing observers to become familiar with a reasonably small number of different classes and students. Familiarity was essential for accurate observational coding, in that observers had to be able to identify individual students readily and be knowledgeable of instructional procedures and materials. Reliability data (discussed below) were obtained by paired observation with one regular observer and one observer who was new to the class. This provided a conservative estimate of reliability and allowed each observer to experience classes other than the eight for which he/she was primarily responsible.

Given responsibility for the observation of eight classes, once each on every other week, an observer generally spent four days a week in classrooms. The fifth day allowed for time to catch up on related data management responsibilities. Hence, the typical observer day started with a school bell, usually at about 8:30 in the morning. Coding began as students entered the classroom door and ended when students walked out of the same door at the end of the day. Observers were responsible for coding all of the time between that entrance and exit for each of six target students. This does not, however, mean that observers coded continuously all day. Most of the detailed coding took place during reading and mathematics instruction, for which specific content categories, student behaviors, and student-teacher

interactions were all recorded. However, during other instructional activities, such as science or art, a single general code was sufficient. During a break period, such as recess or lunch, a single code was also used. Nevertheless, observers often worked hard over recesses or lunch periods, completing identification information on observation pages or reviewing procedures for an interview to be conducted at the end of the school day.

After the last target student left the classroom for the day, observational coding stopped. At that time the observer met briefly with the teacher. The observer asked the teacher if there had been any changes in any of the identified students or in the instructional program for any of them during the past week. The teacher's responses were recorded. On a few occasions a more lengthy interview was conducted. In addition, the teacher's logs for the preceding week were reviewed with the teacher to be sure that no coding problems had arisen. If complete, these logs were collected at this time. After leaving the school, the observer completed a 15-item rating scale describing the classroom and the teacher's behavior on that day. Observers also checked the day's observational coding sheets for incomplete information or procedural errors.

The data collected during the paired observation sessions were used for estimating reliability. The systematic pairing of observers allowed the calculation of inter-observer reliability coefficients for coding instructional time in each of the content categories. Tables 3F-1 through 3F-4 (Appendix 3-F) present indices (Winer, 1962) which indicate the reliability with which one observer codes time for a student on any one day. The numbers entering into the analysis are total time in a given category for a single student during a full day's observation. Data are provided for both the A-B and B-C periods. In almost all cases the coefficients are high. When lower

coefficients occurred, the mean daily times were usually very low. In areas where there was very little recorded time (less than 2 minutes) the reliabilities appear to be unstable. In several of these cases, the coefficients were low (e.g., PV-Grade 5 B-C) but in some cases they were relatively high (e.g., CT-Grade 5 B-C). In content areas where mean daily time was more than two minutes, the correlations are substantial and indicate high agreement between observers on how to categorize reading and mathematics content.

In addition to measures of allocated time, direct observation provided engagement rates and measures of error rate for each student in reading and mathematics. Information on the calculation of the indices and summary statistics for the A-B period are included in Technical Note IV-1a (Filby & Marliave, 1977) and for the B-C period in Technical Note IV-1b (Fisher, Filby, & Marliave, 1977). Additional information on engagement and error rates is presented later in this chapter.

Comparison of allocated time from teacher logs and direct observation system. The teacher log and direct observation procedures provided two independent and very different methods of assessing allocated time. The teacher logs reported data for every day of instruction, whereas, the observation system was based on data from a sample of days (approximately one day per week). As an indirect method of assessing the accuracy of the teacher logs, we compared time in content categories from the teacher logs with the data from direct observation. The reliability for coding of time in content categories was known for the observation system. The comparison was conducted with the assumption that the observation data were probably more accurate than the teacher log data. The observer's sole task was to observe the target students and record information on them continuously throughout the day. Teachers, of course, were primarily involved in teaching and reported

time in content at the end of the school day for their target students. Observers therefore, had a more narrowly defined task and greater opportunity to be accurate. In addition, the observers were more extensively trained on the coding procedures than the teachers (although both groups received considerable training).

To carry out the comparison, a data file was created which included allocated time from observation for the A-B and B-C periods and allocated time from the teacher logs for the days on which observation had been carried out. This file was referred to as the congruency file. Before proceeding, it was necessary to define aggregate content categories wherein the teacher logs and observation would be directly equivalent. The teacher log categories were defined to be relatively narrow. The observational content categories were developed to match the teacher log categories as much as possible. However, the choice of observational content categories was limited by the necessity of having a system which allowed all reading and mathematics content to be recorded without overburdening observers who were required to code many other variables in addition to content. Congruency content categories were defined by aggregating teacher log content categories in order to provide a match with the observation content categories. Appendix 3-G shows the correspondence among congruency, observation, and teacher log content categories.

In examining the congruency data for patterns of discrepancy, some general trends for all teachers in a quadrant² and trends for particular

²The four combinations of subject area and grade level (reading and mathematics at grades 2 and 5) are referred to as "quadrants." For example, in later sections of the report, we refer to the "grade 2 reading quadrant."

teachers were identified. Although miscategorization did occur, there was generally a good match between the logs and observation for the sample as a whole. Appendix 3-G presents correlations between teacher log and observation times for each content category in the congruency file. The correlations were somewhat lower for the A-B period than for the B-C period. This was, to a great extent, due to the fact that the B-C period included approximately three times as much data as the A-B period. In the main, the coefficients represent reasonably high congruency, especially when considered in light of the observation reliability data presented in Appendix F. The reliability coefficients for the observation categories set an upper limit on the degree of congruency that we could hope to obtain between logs and observation.

Measures of allocated and engaged time. To insure consistency in the way allocated time was reported by different teachers, we needed to determine each teacher's accuracy of reporting allocated time. From our experience in Phase III-A and from the congruency file, we knew that some teachers consistently over or underestimate total instructional time. If a teacher overestimated time, adjusting it for engagement would still give the student more time than was actually spent. Consequently, the congruency file was used to calculate an allocation adjustment for each student.

For a given student, total allocated time in the "matched" content for reading or mathematics (RM or MM) as recorded by the observer was divided by the analogous time recorded in the teacher's log. Allocation adjustments, based on the A-C period, were computed separately for reading and mathematics. The grade 2 allocation adjustment ratios ranged from 0.714 to 1.643 in reading and from 0.422 to 1.597 in mathematics. The grade 5 ratios ranged from 0.433 to 2.305 in reading and from 0.504 to 2.727 for mathematics. In each of the four quadrants the means and medians of the distributions of allocation

adjustment ratios were approximately equal to unity. The overall adjustments to the log times were relatively small. A detailed description of the adjustment is presented in Technical Note V-1a, Phase III-B (Marliave, Fisher, & Dishaw, 1977).

Data on student engagement were collected during Phase III-B by direct observation. Individual target students were coded as engaged when they were actively involved in either the substantive content of a reading or mathematics task or learning directions that were required for the performance of substantive aspects of the task. Interim activities, such as sharpening a pencil, getting more paper, or checking off assignments completed, were coded as non-engaged time. Waiting for help and inappropriate off-task behavior were also coded as non-engaged time. Further elaboration of the procedures used to code student engagement are provided in Technical Report I-5 (Marliave, et al., 1977).

Results for the A-B and B-C periods indicate that the average level of student engagement was above 70% at both grades 2 and 5 (Filby & Marliave, 1977; Fisher, Filby, & Marliave, 1977). It should be noted that this high level of engagement does not necessarily reflect a high level of efficiency. The rate of observed engagement reflects only the proportion of allocated reading and mathematics time during which students were actively involved in their tasks. Allocated reading and mathematics time were defined to exclude general management activities, time required for transitions from one activity to another, and "down time" during which no activity was provided for a given student. These three categories of time accounted for an average of 44 minutes per school day at grade 2 and 47 minutes per day at grade 5 (Fisher, Filby, & Marliave, 1977). Presumably, much of this time detracted from time that might otherwise have been allocated to reading or mathematics.

The adjustment of teacher logs for allocation differences (discussed

above) resulted in log records of allocated time in reading and mathematics to which the observed rates of engagement could be appropriately applied. The reliability of the observational data on engagement was examined before applying these ratios to the adjusted allocated time from the teacher logs. Appendix 3-H provides information on the reliability of the engagement variables. Additional discussion on engagement is presented in Technical Notes V-1a and V-2a (Marliave, Fisher, & Dishaw, 1977, 1978).

After considering the reliability information, observed engagement rates for individual students were computed in each of reading and mathematics for the entire A-C period. These coefficients were used to adjust both A-B and B-C allocated time data from teacher logs to provide measures of engaged time.

The engagement rate adjustment coefficients were based on the "matched" reading and mathematics content categories (RM and MM). This insured that the adjustments corresponded to the specific content categories of most interest to the study. The actual computation of these engagement rates was carried out by simply taking the ratio of total engaged time observed for matched reading or mathematics over total allocated time observed for the same content. The teacher log data for specific content categories were adjusted for observed engagement by taking the product of the log time for a student and that student's observed ratio of engaged to allocated time.

Comparison of sources of time data. The data collected during Phase III-B provided a choice among three alternative sources of data on instructional time: teacher logs (allocated time), teacher logs adjusted according to observed allocation and engagement ratios (as described in the previous section), and observational coding of engaged time. A series of regression analyses was conducted to provide an empirical basis for comparing the three alternative sources of time data. Sixteen sets of regressions were used, two

for each of the eight grade by period by reading/mathematics data sets. Each of these regressions followed the simple model of post-achievement scores regressed on the corresponding pre-achievement scores and instructional time in a corresponding content category. Each of the 16 regressions was repeated three times: once using each of the three alternative sources of time data (yielding a total of 48 regressions). The results were examined to compare the relative predictive power of the three different data sources for instructional time as predictors of student post achievement adjusted for pre achievement. The detailed results of these regressions are included in Technical Note V-1a (Marliave, Fisher, & Dishaw, 1977).

Generally, it was concluded that the adjusted teacher logs provided the best prediction of student learning. However, the three alternative time sources were much more similar in their predictive efficacy than had been expected. Of particular note was the finding that unadjusted logs were almost as potent in these regressions as were adjusted logs.

As a result of these analyses, we chose the teacher logs adjusted for engagement as the time source for preliminary examination of the relationship between instructional time and student achievement. In addition to these analyses using teacher logs adjusted for engagement as a time source, we also report analyses wherein allocated time (teacher logs adjusted with the allocation ratio) and student engagement ratios (from direct observation) are analyzed as separate variables.

Student error rate. Data on error rates for individual students were collected during Phase III-B by direct observation. Error rates were coded relative to the individual target student's ability to perform instructional tasks. That is, error rates were not assigned to the tasks themselves. Rather, the error rate coded was a function of both the student and the

particular task on which the student was working.

Three levels of error rate were coded: low, medium, and high. The low level was coded when the student was able to perform a task with no errors except those that could be attributed to chance (careless errors). The high level was coded when the student was not able to respond correctly, except by chance (guessing). Medium error rate was coded for everything in between low and high (some non-chance errors and some non-chance correct responses). Technical Report I-5 (Marliave, et al., 1977) provides further elaboration of the procedures used to code error rate.

Results for the A-B and B-C periods (Filby & Marliave, 1977; Fisher, Filby, & Marliave, 1977) indicate that the average proportion of time spent by students at each of the low and medium error rates was approximately 50%. The average proportion of time spent at the high level was about 2%. These proportions indicate that students spent a considerable amount of time reviewing and practicing what they had learned (as indicated by the high proportion of time with low error rate) and very little time working on tasks that they were unable to perform at all (as indicated by the very small proportion of time with high error rate).

We decided that observed error rate variables for analyses with achievement should be based on the entire A-C data set so that they would include a greater number of occasions. This decision was based upon analysis of the reliability of the error rate variables (see Appendix 3-I). It was expected that levels of error rate for individual students might be generally stable from the A-B period to the B-C period, insofar as they reflect the skill of the teacher in diagnosing student instructional needs or the ability of the student relative to other students in the class. Furthermore, A-C error rate variables should improve upon the B-C reliabilities. Although the

B-C reliabilities were generally acceptable, those for the high error rate variable were approximately .50, with one case (grade 2 reading matched) at about .20 (see Appendix I). We expected that sampling error would be especially serious for the infrequently occurring "high" category, but it is anticipated that this variable may be particularly potent where it does occur. The use of A-C difficulty variables should be especially advantageous for the "high" category.

The A-C error rate variables were used in both the A-B and B-C analyses of ALT and achievement. For analysis purposes, the error rate variables were computed as proportions of total time in reading content (or in mathematics content). This computation consisted of taking the ratio of total reading or mathematics time at one of the three levels of error rate over total allocated time for the same content (reading or mathematics).

Several error rate variables may be defined using these proportions. First, the simple proportions of low, medium, or high time may be analyzed. In addition, it would be reasonable to enter two of the variables as separate predictors. Another alternative is the product of two error rates, such as the proportion medium and the proportion low. This would give maximum weight to a balance between the medium and low error rates (50% of each) and minimum weight to extremely large proportions of either one (100% of either medium or low would result in a product of zero). One additional error rate variable meriting consideration is the product of an error rate proportion and engaged time in the relevant content area. Taking the product of the error rate and engaged time yields an estimate of engaged time at that level of error rate, which more nearly conforms to the ALT concept than the separate entry into regressions of engaged time as one variable and an error rate as another variable. These various alternatives were explored in the analyses of A-B

period data, as presented and discussed in Chapter 4.

Teaching Process Measures

Interactive teaching behavior. Interactive teaching behavior was assessed along with ALT in the direct observation system. Each time an event was sampled for a target student, the observer noted whether or not any instructional behavior was being directed toward the student at that moment. If some instructional behavior was occurring, the observer coded the nature of the behavior, and also the source and focus of the behavior. The observer also coded instructional setting -- self-paced or other-paced. See Appendix 3-E for a description of coding categories in the observation system.

The teaching behaviors included in the observation system can be grouped according to the three functions of interactive teaching behavior (presentation, monitoring, and feedback). Three types of presentation are included -- explanation-planned, explanation-need, and structure/direct. In planned explanation, the teacher (or other source) provides substantive academic information to the student(s). The presentation is a normal part of the curriculum, like a lecture. In explanation based on need, the teacher provides substantive information specifically in response to perceived needs or misunderstandings of the student(s). Structure/direct involves presentation of procedural information. The teacher may be creating a framework of expectation for the lesson ("Now we are going to do our math speed test.") or may be making a specific request ("Get out your reading book.")

Two types of monitoring are included in the system -- academic observation and academic question. Monitoring is defined as behavior which provides information for the teacher about how well a student can perform a particular task. In academic observation, some evidence of performance is overtly

present, and the teacher is examining the evidence. For instance, the teacher might look at a student's paper while he does seatwork. In academic questioning, the teacher asks a student a question in order to elicit an observable response.

Two types of feedback are included -- academic feedback and task engagement feedback. Feedback lets the student know how he is doing; it helps him to evaluate his performance. In academic feedback, the student receives information about his academic performance, whether his answer is right or wrong or what the right answer is. In task engagement feedback, the student is told whether his behavior is acceptable or unacceptable, or what he should be doing.

It should be noted that the BTES observation system reflects the instruction received by an individual student. The system generates information about how much time an individual student spends receiving particular behaviors. The system does not provide a complete record of the teacher's time.

The interactive teaching behaviors are described in more detail in Chapter 5, where the relationships between interactive teaching behavior and student engagement rate are examined. Chapter 5 includes interobserver reliability estimates, descriptive statistics on the frequency of occurrence of different interactive teaching behaviors within setting, and intercorrelations of teaching behaviors.

Measures of planning and general characteristics. In addition to the time-based observation of interactive teaching behaviors, a number of other methods were used to obtain information about the teacher and about the classroom instructional program. Interviews were used to elicit information about the planning functions -- diagnosis and prescription. Ratings were

made at the end of each observation day to assess qualitative aspects of the instructional program and the classroom climate. In addition, teachers responded to an aptitude test and an attitude questionnaire.

The total number of measures obtained was larger than the number of classes in the study, a mind-boggling impossibility for quantitative analysis. This was done because we anticipated that not all measures would prove reliable and because we wanted multiple opportunities to measure important variables. Nevertheless, it was a major undertaking to reduce the size of the variable pool in a manner that was satisfying to both the logic of the measures and the empirical evidence of interrelationships.

The following sections of this chapter will describe briefly the different measures that were collected and the process of variable selection and definition. Further details can be found in Technical Note V-1b (Filby & Cahen, 1977). The instruments themselves, and instruction for their use, are presented in Technical Report I-5 (Marliave, Fisher, Filby, & Dishaw, 1977).

Initial measures -- interviews. In the fall, an initial interview and planning interview were conducted with each teacher. As part of the planning interview, teachers were asked to predict which test items from a representative set of items each target student would answer correctly. Teachers were also asked to describe the strengths and weaknesses of each student and the instructional plans for each student.

In addition to the fall interviews, a brief interview was conducted on each observation day. The teacher was asked to describe any changes that had occurred in the target students or in their instructional programs.

A follow-up interview was conducted in the spring. Teachers were again asked to predict student test performance. The items chosen for this prediction task were items which some but not all students would master by

the end of the year. They covered a range of content areas.

A variety of measures were derived from the accumulated wealth of interview information. In the spring, the field workers reviewed the interview information and used this information to perform a number of ratings. Some ratings covered aspects of diagnosis and prescription; for instance, the degree to which the teacher noted individual characteristics of students ("differentiated perceptions"), the specificity of goals stated by the teacher, and the flexibility with which curricula were implemented. Other ratings dealt with the general value orientation of the teacher, for instance, the value placed on affective outcomes. All ratings were completed by each of the two field workers regularly assigned to each classroom. The score assigned to each teacher was the average of the two field workers.

In addition to these ratings, several more direct measures were derived. One set of measures dealt with item prediction. Three different item prediction measures were created in each subject-matter area. First, a count was made of the number of times the teacher correctly predicted how a student would do on an item (pass or fail). Separate scores were calculated for fall and spring. With four items and six students, this amounted to a 24 item "test." In addition to this scoring procedure, an item-prediction discrepancy score was calculated for each student. This discrepancy score was the absolute value of the difference between the number of items the student answered correctly, including fall and spring, and the number of items the teacher said the student would answer correctly. This procedure was intended to minimize chance fluctuations in whether or not a student answered a single item correctly.

Besides these item prediction scores, another set of measures came

directly from interviews. Each teacher's weekly interview protocols were analyzed and coded. Most often, teachers reported no change in students or programs. This was particularly true in grade five. When teachers did comment, each statement was coded as one of three types -- diagnosis, program change-based-on-need, or general program change. A diagnostic statement was defined as any statement about a student's academic performance. Such a statement indicated that the teacher noticed something about a student's strengths, or weaknesses. For instance, a teacher might comment that a student was having a hard time understanding fractions or that a student was finishing his work more quickly now. A program-change-based-on-need statement reported a situation where the teacher acted on diagnosis and made a change in the student's program based on perceived students needs. For instance, if a student showed difficulty understanding a concept, the teacher might provide special tutoring or assign extra homework. The third type of statement was a general program change. This statement reflected a situation where program changes took place as part of the normal progression of curriculum. One common example of this situation is finishing one reading book and moving on to the next book in the series.

All weekly interviews were coded by two coders. Special problems with coding were brought to a staff member for arbitration. Scores were created by summing for each student the number of occurrences of each type of statement. Following reliability analysis, scores were averaged across coders.

Initial measures -- daily rating scales. At the end of each day of observation, the field worker rated the classroom and the instructional program on 15 scales. These scales were called: cooperation, student responsibility, conviviality, acceptance of feelings, cognitive task orientation, attention to students, expectation of student success, knowledge

of reading pedagogy, knowledge of math pedagogy, clarity, abruptness, flexibility, complexity of communication, consistency of message, and structuring.

Since observation took place approximately once each week, there were about 5 ratings on each scale over the A-B period, and 15 ratings over the B-C period. These ratings were averaged to get an A-B average score and a B-C average score on each scale.

Initial measures -- aptitude and attitude. In addition to measuring characteristics of the ongoing instructional program, we also included a few measures of teacher characteristics. One such measure was a type of aptitude score. With the focus on diagnosis and prescription in the BTES model, an aptitude of hypothesized importance is knowledge of subject matter. A teacher who knows basic concepts and terminology in reading and mathematics pedagogy should be more likely to diagnose individual student strengths and weaknesses and plan an appropriate instructional program.

Knowledge of subject matter was measured by a content coding test given in conjunction with training in log keeping procedures. All teachers in the study participated in a one-day training session to learn how to keep records of allocated time. Log keeping required each teacher to identify the content being covered in daily lessons and workbook pages. In the log manual and in discussions during the training session, terms were defined and examples were provided. Following training, each teacher was asked to code the content of a set of representative workbook pages. Each teacher's score was the total number of correct content codes.

Three attitude measures were also obtained on each teacher (see Appendix D in Howell & Rice, 1977). In the spring, each teacher filled out a questionnaire covering attitude toward reading and attitude toward mathematics. Other items in the same questionnaire dealt with aspirations,

including such items as "I would enjoy leading a workshop in teaching techniques" and "I would be interested in acquiring a supervisory certificate." All items were rated on a 1 to 5 scale, with 5 representing strong agreement (a positive attitude). Three scale scores were created by adding the ratings on component items. The scales were independent and showed reasonably high internal consistency.

Process of variable selection and definition. We were faced with the task of reducing our large set of planning and general measures to a smaller set. First, we had to understand the interrelationships among measures. When decisions were made to combine measures, we had to select a procedure to combine single measures into a cluster score. We used an approach that included both logical and statistical inputs. The goal of clustering, in addition to reducing the number of independent variables, was to produce sets of variables with the following qualities:

1. Measures within a set would be homogeneous:
2. Sets would be as independent as possible of other sets:
3. Sets would be as free as possible of extraneous sources of variance (such as rating "halo" or observer bias, method variance, etc.):
4. The measures in a set would be logically related (a form of construct validation), and be meaningfully represented by a descriptive label.

This last criterion was considered especially important. Before dimensionality analyses were undertaken, we knew that our model of instruction should be used to guide our dimensionality investigation. We attempted to use the teaching functions, particularly diagnosis and prescription, as a "map" for clustering our measures.

The actual process of dimensionality analysis, variable definition and selection proceeded in stages. In the first stage, the daily rating

measures were examined for interrelationships. Clusters of daily rating scales were created based on this analysis. In the second stage, the interview-based measures were examined. Again some combinations were created. At this point, we also looked at partial correlations and created some residual score variables. Finally, this set of intermediate measures from all sources was subjected to a final analysis of interrelationships. Decisions were made about final combinations of measures and about which variables to use for analysis with ALT outcomes and with achievement. Detailed descriptions of these procedures are included in Technical Note V-1b (Filby & Cahen, 1977).

Through this process, three diagnosis variables, three prescription variables, and five general variables were selected. The diagnosis and prescription variables are described in Chapter 6, where they are related to student error rate. Reliabilities and descriptive statistics are also provided in Chapter 6. The general variables are introduced in Chapter 5. Again, variables are described in some detail, reliabilities and descriptive statistics are provided.

CHAPTER 4: ACADEMIC LEARNING TIME AND STUDENT ACHIEVEMENT

Academic Learning Time (ALT) comprises three dimensions that represent ongoing student learning in the time metric. The first dimension is the relevance of the student's instructional activity to a given academic outcome. This dimension is measured in terms of teacher allocations of instructional time to particular content categories. The other two dimensions characterize the manner in which the student undertakes (or fails to undertake) a given instructional activity. They are student engagement in the activity and student error rate in responses required for that activity.

This chapter presents brief descriptive statistics on each of the three ALT components and the results of multiple regression analyses showing relationships between these three components and student learning (the latter being represented by pre and post achievement test scores). Both descriptive and analytic statistics are presented for each of the three components separately. There is some discussion of attempts to combine certain components of ALT, but these attempts were generally not advantageous. It should be underscored, however, that ALT as represented in the model (see Chapter 2) is the simultaneous occurrence of these three dimensions. Limitations in the reliability of measurement procedures (see Chapter 3) precluded the analysis of time for this simultaneous occurrence. That is, it was possible to measure allocated time in a specific content category, but not to measure engaged time at a low error rate within a specific content category. Rather, engagement and error rate data were aggregated to the total reading and total mathematics level (and even to the total A-C period level), in order to enhance reliability. Hence, the ALT components were measured as separate

variables. As noted above, attempts to combine them (by taking products) showed no advantage over analyses of the separate components. Therefore, the data presented here are not an ideal representation of ALT, but rather a practical approximation.

It should also be noted that analyses of teaching behaviors in relationship to ALT examine engagement and error rate as separate variables, differentially influenced by the teaching behaviors (see Chapters 5 and 6). This is in accordance with the ALT model. It would be unreasonable to expect that a given teaching behavior will influence all components of ALT in the same manner and with equal strength. Therefore, the validation of each ALT dimension separately, as presented in this chapter, is appropriate for the analyses of teaching behaviors that will follow.

The ALT Variables

The three dimensions of ALT (time allocation, engagement rate, and error rate) will be reviewed briefly before descriptive statistics and regression analyses are presented. A more detailed explication of these three dimensions is included in the manual used for training observers (Marliave, Fisher, Filby, & Dishaw, 1977).

Time allocation refers to the amount of time in a particular activity designated by the teacher for a given student. This was coded without regard to whether or not the student actually attended to the activity. Therefore, a teacher could allocate 20 minutes of spelling instruction for the entire class. One student might spend the entire 20 minutes working on spelling while another student talks with his/her neighbor and doddles on an eraser. Despite these different student behaviors, the time allocation is the same for both students (20 minutes of spelling).

Engagement rate refers to the proportion of total allocated time in reading or mathematics during which a given student actually attended to the designated activity. Therefore, in the example above, the first student had an engagement rate of 100% during the spelling activity, while the second student's engagement rate was zero%. The engagement rate used in analyses, however, would be based on an average rate of student engagement over all the days of reading or mathematics instruction observed.

A student was considered to be engaged in an activity whenever he/she was involved in the academic substance of that activity. That is, writing answers to problems, reading aloud or silently, and listening to the teacher's explanation were all coded as engagement in the reading or mathematics activity. However, time spent sharpening pencils, getting more paper, or sitting and waiting for help was coded as non-engaged, because none of these behaviors involve the academic substance of the reading or mathematics activity. That is, a student may have to sharpen his/her pencil in order to complete a mathematics worksheet, but time spent sharpening pencils is not directly related to learning the mathematics content of the worksheet. Of course, student behaviors that are totally inappropriate to the instructional activity, such as socializing or daydreaming, were also coded as non-engaged.

Error rate refers to the proportion of incorrect responses made by the student during an instructional task. Three categories of error rate were coded (low, medium, and high). Low error was coded when the student made no errors, other than careless errors (attributable to chance). High error was coded when the student made no correct responses (all responses were errors), except by guessing (correct responses attributable to chance). Medium error included anything between low and high error. The purpose of

this three category schema was to distinguish between tasks where the student was practicing or reviewing material that he/she already understood (low error), tasks where the student was totally lost and unable to respond appropriately (high error), and tasks where the student had partial understanding of the material to be learned (medium error). Note that for most students the high error category would occur infrequently. As will be discussed later in this chapter, it was found that the high error category accounted for only about three percent of reading and mathematics instruction, while the low and medium error categories each accounted for about half of the remaining instruction.

In order to code a student's error rate for a particular task, the observer had to actually examine several of that student's responses, counting the number of correct and incorrect student answers. It was also necessary to consider the likelihood of responding correctly by guessing. That is, a student should be able to guess correctly on about half of 100 true-false questions, so that 50 errors on such a task would be coded as a high error rate. However, a series of objectively scored open-ended questions, such as dictated spelling questions, might have a zero probability of correct response by outright guessing. In such a case, correct responses on even a quarter of the questions would be coded as medium error.

It is important to note that the coding of error rate was based on student responses to a task, not on the task per se. Therefore, the same task could be coded as low error for one student and high error for another student in the same class. For example, two fifth grade student might be drilling each other in basic multiplication facts using flash cards. One student might have already learned these basic facts so that the task is a low error review for him/her. The other student may be learning the basic

facts for the first time. Therefore, the task would probably be medium error for that student, in that he/she might understand the concept of multiplication well enough to know the most simple basic facts (2 times 2 equals 4) and to learn others without too much trouble. However, this student might be completely confused, have no concept of multiplication, and try to apply addition or subtraction principles to the task. In this case, the student might respond correctly only by guessing, so that high error would be coded.

It should be clear that error rate can be coded only on the basis of overt responses. However, students often work on tasks that involve no immediate overt responses. Such tasks include silent reading and listening to the teacher explain a mathematics operation. In addition, students may sometimes respond as a group, so that no individual student's ability to respond correctly can be determined accurately. In these cases, error rate was coded on the basis of preceding or subsequent overt responses on the same kind of instructional task. For example, a student might read along silently while another student reads aloud in a reading group. However, it is likely that the student reading silently would eventually take a turn at reading aloud. That student's error rate during his/her turn reading aloud would be coded for the previous silent reading by the same student. Similarly, a teacher will often explain a mathematics operation after students have had some trouble carrying out that operation, or just before the students begin an assignment involving the operation. Overt student responses during the preceding or subsequent task would be coded for the time during which students listened to the teacher's explanation.

In some cases, no overt response related to a given task might be observed during an instructional period. For example, a student might spend an entire period reading silently from a book that is not similar to

other materials the observer has seen this student read recently. In such a case, the observer would wait until the next recess and ask the student to read a paragraph or two from the book. Error rate would be coded on the basis of this oral reading requested by the observer. Observers avoided directly requesting student responses except when this was necessary for the accurate coding of error rate.

Descriptive Data on ALT

General time allocations. Table 4-1 presents an overview of the time allocations by teachers in this study. Means and standard deviations are presented for observational data from grades 2 and 5 during the A-B and B-C periods. Note that the time allocations are quite consistent across the two periods (within grade). Note also that the additional hour of instruction at grade 5 is devoted largely to reading and language arts and to other academic instruction (such as science and social studies). Relatively little time is spent on management (milk money distribution, etc.) and even less time is without any designated activity (the "wait" category). However, more than half an hour is expended on transitions, primarily movement from one activity to another, at both grades 2 and 5. Apparently students do not learn to conduct transitions more efficiently as they grow older.

The last four general time categories in Table 4-1 (break, wait, transition, and management) represent non-instructional activities. Appendix 4-A presents a more complete discussion of these non-instructional activities and analyses pertaining to them.

Specific time allocations. Table 4-2 presents means and standard deviations for allocated time in specific reading mathematics content categories, as recorded in grade 2 teacher logs. Table 4-3 presents the corresponding data for grade 5. Data are presented for both total minutes

Table 4-1

Where the Day Goes --
Minutes per Day in General Content Categories

Time Category	Grade 2				Grade 5			
	A-B Period		B-C Period		A-B Period		B-C Period	
Reading and Language Arts	85	(21)	88	(16)	113	(24)	110	(20)
Mathematics	37	(12)	36	(8)	41	(13)	44	(13)
Other Academic	9	(10)	8	(6)	25	(18)	17	(11)
Non-Academic	54	(22)	55	(17)	52	(29)	65	(18)
Break	76	(10)	75	(9)	80	(11)	77	(10)
Wait	4	(4)	4	(3)	5	(4)	4	(3)
Transition	34	(10)	34	(6)	32	(8)	34	(7)
Management	7	(4)	6	(3)	9	(5)	9	(5)
Length of the School Day	306	(14)	307	(10)	356	(19)	361	(13)

Entries are: Mean (Standard Deviation)
Data are based on observation.

Table 4-2

Descriptive Data
Academic Learning Time
Time Allocations in Specific Content Categories (A-C Period)
Reading and Mathematics, Grade 2 (N=139)

Content Category	Student Average Total Minutes, A-C Period (99 Days)	Estimated ^a Student Average Minutes Per Day, A-C Period
Reading		
Total Comprehension	964 (383)	9.8
Decoding Blends and Long Vowels	489 (331)	5.0
Decoding Variant Consonants	72 (94)	0.7
Spelling	527 (378)	5.4
Word Structure - Meaningful Units	307 (206)	3.1
Word Structure - Syllables	80 (102)	0.8
Total Reading	5082 (1401)	59.8
Mathematics		
Addition and Subtraction- No Regrouping	729 (369)	7.4
Addition and Subtraction- Speeded Tests	69 (84)	0.7
Addition and Subtraction- with Regrouping-Algorithms	333 (296)	3.4
Computational Transfer	559 (374)	5.7
Place Value and Numerals	427 (218)	4.3
Word Problems	169 (159)	1.7
Money	115 (110)	1.2
Fractions	71 (89)	0.7
Total Mathematics	2967 (692)	30.0

(Standard deviations for total minutes are in parentheses.)
Data are based on teacher logs.

^a This estimate is based on the average total minutes per student for the period, divided by the average number of school days each student was present during the period.

Table 4-3
 Descriptive Data
 Academic Learning Time
 Time Allocations in Specific Content Categories (A-C Period)
 Reading and Mathematics, Grade 5 (N=122)

Content Category	Student Average Total Minutes, A-C Period (95 Days)	Estimated ^a Student Average Minutes Per Day, A-C Period
Reading		
Total Word Meaning	822 (270)	8.7
Total Comprehension	2091 (1214)	22.1
Word Structure - Syllables	151 (109)	1.6
Total Reading	5838 (1481)	61.7
Mathematics		
Total Geometry	187 (233)	2.0
Total Multiplication	560 (500)	5.9
Multiplication Basic Facts	70 (139)	0.7
Division	740 (507)	7.8
Fractions	905 (815)	9.6
Computational Transfer	220 (199)	2.3
Word Problems	216 (257)	2.3
Total Mathematics	3218 (665)	34.0

(Standard deviations for total minutes are in parentheses.)
 Data are based on teacher logs.

^a This estimate is based on the average total minutes per student for the period, divided by the average number of school days each student was present during the period.

of the entire A-C period of the study and for the average daily minutes. Note that the average daily times in total reading and total mathematics displayed here are substantially less than that shown in Table 4-1. This is an artifact of definition. The total reading and mathematics times reported in Tables 4-2 and 4-3 exclude developmental and related instructional activities for which no achievement test scales were administered. Therefore, these total times correspond directly to the total reading and mathematics achievement batteries used. It was total reading and mathematics time as defined for Tables 4-2 and 4-3 that were analyzed in relation to total reading and mathematics achievement.

Table 4-4 presents means and standard deviations for engagement rate, low error rate, and high error rate, as observed over the entire A-C period of the study. Note that the average engagement rate for the A-C period was between 70 and 75 percent for both grades (2 and 5) and both content domains (reading and mathematics). One reason for these relatively high rates of engagement was the exclusion of transitional activities from allocated reading and mathematics time as recorded for this study. That is, time during which students were moving from one activity to another, or simply waiting for the next activity to begin, was excluded from the allocated reading and mathematics categories. This time might otherwise have been recorded as non-engaged reading and mathematics time. Engagement rates would have been much lower if transitional activities had been coded as non-engaged time. As discussed previously, teacher records of total reading and mathematics allocations were corrected according to observer allocation data. This correction assured that the exclusion of transitional activities from reading and mathematics allocations, for which the observers were carefully trained, was consistently applied to the teacher records of allocated time.

Table 4-4

Descriptive Data
Academic Learning Time
Engagement and Error Rate (A-C Period)

Grade	Content Domain	Proportion Engaged ^a	Proportion Low Error Rate	Proportion High Error Rate
² (N=139)	Reading	.74 (.10)	.50 (.18)	.02 (.03)
² (N=139)	Mathematics	.70 (.12)	.50 (.21)	.05 (.07)
⁵ (N=122)	Reading	.75 (.11)	.46 (.15)	.01 (.02)
⁵ (N=122)	Mathematics	.73 (.13)	.34 (.21)	.03 (.06)

Entries are: Mean (Standard Deviation)
Data are based on observation.

^aNote that engagement rates were coded for reading and mathematics, but not for transition and wait time between instructional activities. This resulted in higher engagement rates than would otherwise have been obtained.

Table 4-4 also indicates that the high error rate category accounts for a very small proportion of instructional time, ranging from one to five percent. The low error rate category accounts for one-third to one-half of the remaining time. All other instructional time in reading and mathematics was coded as medium error rate (the only other category possible). Therefore, the low and medium error rates account for the vast majority of instructional time, so that these two categories are essentially ipsative. For this reason, these two categories were never both entered as separate variables in the same analysis.

Intercorrelations of ALT variables. Table 4-5 presents intercorrelations among allocated time, engagement rate, low error rate, and high error rate. Allocated time is included for both the A-B and the B-C periods, while engagement rate, low error rate, and high error rate are included for the A-C period. Correlations are presented for each of the four quadrants (grade 2 reading, grade 2 mathematics, grade 5 reading, and grade 5 mathematics). Significant correlations are starred ($p \leq .10$).

The correlations between allocated time and the other three variables (engagement rate, low error rate, and high error rate) show no consistent pattern overall. No significant correlations were found in grade 2 reading or grade 5 mathematics. Allocated time in grade 2 mathematics, for both the A-B and B-C periods, was negatively correlated with both engagement rate and low error rate ($p \leq .10$). Allocated time in grade 5 reading for the B-C period was positively correlated with engagement rate ($p \leq .10$). Allocated time in grade 5 reading for the A-B period was negatively correlated with both low error rate and high error rate ($p \leq .10$).

This set of correlations does not lend itself to any simple interpretation. Any generalizations based on these results must involve a fair degree of speculation. However, the results for grade 2 mathematics may indicate

Table 4-5
ALT Variable Intercorrelations

Intercorrelations at Grade 2 and Grade 5 between:
Allocated Time (Total Reading and Total Mathematics for the AB and BC Periods), Engagement Rate (AC Period),
Low Error Rate (AC Period), and High Error Rate (AC Period)

Grade 2 (AB Period N=127; BC and AC Periods N=139)				Grade 5 (N=122)		
	3. Engagement Rate (Reading or Mathematics)	4. Low Error Rate (Reading or Mathematics)	5. High Error Rate (Reading or Mathematics)	3. Engagement Rate (Reading or Mathematics)	4. Low Error Rate (Reading or Mathematics)	5. High Error Rate (Reading or Mathematics)
1. Allocated Time Total Reading AB Period	-.05	.05	-.01	.10	-.25**	-.38**
2. Allocated Time Total Reading BC Period	.05	-.04	-.09	.26**	.09	-.14
3. Engagement Rate Total Reading AC Period	1.00	.13	-.36**	1.00	.11	-.20**
4. Low Error Rate Total Reading AC Period		1.00	-.37**		1.00	.08
5. High Error Rate Total Reading AC Period			1.00			1.00
1. Allocated Time Total Mathematics AB Period	-.25**	-.17*	.14	-.12	-.05	-.11
2. Allocated Time Total Mathematics BC Period	-.28**	-.25**	.04	.03	-.01	.06
3. Engagement Rate Total Mathematics AC Period	1.00	.28**	-.33**	1.00	.15*	-.08
4. Low Error Rate Total Mathematics AC Period		1.00	-.38**		1.00	-.16*
5. High Error Rate Total Mathematics AC Period			1.00			1.00

4-13

82

that those teachers who allocated relatively large amounts of time in this area also tended to provide relatively little work that students could complete at a low error rate. This, in turn, could account for the negative association of time allocation and engagement rate in grade 2 mathematics, insofar as greater time allocations at medium and high error rates would be associated with lower levels of student engagement (see discussion below).

In grade 5 reading, time allocation is positively associated with engagement rates, at least during the B-C period. However, time allocations in this quadrant are negatively associated with high error rate (particularly for the A-B period) and show a negative association with low error rate only for the A-B period (where the correlation with engagement was nonsignificant).

One could infer, then, that relatively high levels of time allocation are associated with lower levels of engagement only when the additional allocations result in medium or high student error rates (as in grade 2 mathematics). That is, there does not appear to be any consistent pattern of lower student engagement when students are asked to work on reading or mathematics for relatively long periods of time. Rather, it appears that student engagement decreases when additional time is allocated on relatively more difficult activities (as for grade 2 mathematics), but may increase when the additional time is on relatively less difficult activities (as for grade 5 reading during B-C).

The results for the intercorrelations of engagement and error rates (Table 4-5) are reasonably consistent across grade and subject matter. In addition, these analyses tend to support the discussion of the allocated time correlations, above. Engagement rate is positively correlated with low error rate and negatively correlated with high error rate in all four quadrants, where two of the positive and three of the negative correlations

are significant ($p \leq .10$). These findings may indicate that students pay more attention when they are working on tasks on which they commit relatively few errors. That is, students are more easily distracted when they have greater difficulty completing a task correctly. Note, however, that these correlative findings could also be interpreted as showing that when students pay more attention, then they make fewer errors. Both interpretations probably have some validity.

Table 4-5 also shows that low error rate and high error rate are negatively correlated ($p \leq .10$) in all quadrants except grade 5 reading (where the correlation is nonsignificant, positive). This finding is not surprising, in that low and high error rates are conceptually opposite. The predominance of one would suggest an infrequency of the other (in relative terms).

Relationships between ALT and Student Learning

Analyses of the relationship between Academic Learning Time and student achievement were conducted on two sets of data: 1. the A-B period data from mid-October through early December, and 2. the B-C period data from early January through the end of April. These two data sets were intended, for the most part, to provide replications of analyses. In practice, the A-B analyses served a more exploratory function, while the B-C analyses were more confirmatory. For example, the A-B analyses were used to explore the various possible relationships between student error rate on instructional tasks and student learning outcomes. The B-C analyses provided confirmation of the error rate effects found in A-B analyses. The B-C analyses are probably more definitive and trustworthy than the A-B analyses. This is due

in part to the replicative, confirmatory nature of the B-C analyses, and in part, to the fact that the B-C data set was more than twice as large as the A-B data set. However, the relationships that received general support from both sets of analyses are those that deserve the greatest attention, elaboration, and possible application.

A-B analyses. Table 4-6 presents the results of regressions analyzing the effects of engaged time in reading and mathematics at grades 2 and 5 during the A-B period. Engaged time is the product of allocated time (based on teacher logs) and engagement rate (from observation). Results are given for the subcategories of reading and mathematics that were tested, as well as total reading and total mathematics. The proportion of residual variance in the post achievement measure that was accounted for by engaged time is reported for each content category. Residual variance, in this report, refers to the variance remaining in the post achievement measure after pre achievement has been "removed" through regression. This residual variance may be thought of as a rough estimate of student "learning," where learning is the difference between a student's actual post achievement score and that which one would have expected knowing only his/her pre achievement score. This difference may be attributed to events that occurred between the two tests, presumably positive and/or negative learning events.

The proportion of residual variance accounted for by engaged time is given only when the significance of that effect was at or below the .10 level. Significance at or below the .01 level is also denoted. In addition, the sign of the effect is given, where positive effects indicate that the actual post achievement score was greater than expected on the basis of pre achievement, while negative effects indicate the reverse. The proportion of total post achievement variance accounted for by pre achievement is also given. The difference between this proportion and 1.00 is the residual variance (as a

Table 4-6

Regressions Analyzing the Effects of Engaged Time

Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the A-B Period, Regressing Post Achievement on Pre Achievement and Engaged Time

Grade 2 (N=127)					Grade 5 (N=122)				
Content Category for Post Achievement, Pre Achievement and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Engaged Time ^a Effects			Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Engaged Time ^a Effects		
		Significance * p $\leq$.10 ** p $\leq$.01	Sign + positive - negative	Proportion of Residual ^b Variance			Significance * p $\leq$.10 ** p $\leq$.01	Sign + positive - negative	Proportion of Residual ^b Variance
Total Comprehension	.31	**	+	.08	Total Word Meaning	.58			
Total Decoding	.65				Total Comprehension	.71			
Total Word Structure	.40	*	+	.04	Word Structure Syllables	.41			
Total Reading	.71	*	+	.05	Total Reading	.81			
Addition and Subtraction No Regrouping	.41				Total Geometry	.24	**	+	.13
Addition and Subtraction Speeded Tests	.49				Total Multiplication	.28	**	+	.14
Computational Transfer	.28				Multiplication Speeded Tests, Basic Facts Time	.54	*	-	.03
Place Value and Numerals	.38				Division	.24	**	+	.06
Word Problems	.29				Fractions	.08	**	+	.20
Money	.43				Computational Transfer	.55			
Total Mathematics	.63				Word Problems	.22			
					Total Mathematics	.70			

^aComputed by taking the product of Allocated Time in the content category (based on teacher logs) and Engagement Rate (based on observation).^bThis is the proportion of the residual variance accounted for by Engaged Time, where residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed."

proportion of total post achievement variance).

The results in Table 4-6 indicate that engaged time in the A-B analyses was positively related to student learning. In the strongest case, 20 per cent of the residual variance was accounted for by engaged time (fractions at grade 5). Note also that all but one of the significant ($p \leq .10$) engaged time effects were positive, and the one negative effect (multiplication speeded test at grade 5) was the smallest of the significant effects. It should also be noted, however, that no significant effects were found in grade 2 mathematics or grade 5 reading for the A-B period data on engaged time.

The A-B analyses of engaged time were followed by a series of analyses of error rate. The first two sets of regressions examined medium error rate and high error rate. The analyses of medium error rate showed a surprising negative relationship with learning outcomes. This was consistent for grade 2 mathematics and grade 5 reading. High error rate was found to have negative effects in grade 2 reading and grade 5 mathematics. These negative effects were expected for high error rate. (See Technical Note V-1a, Marliave, Fisher, and Dishaw, 1977, for a more complete account of these analyses.)

As an exploration of the unexpected negative effects for medium error rate, the product of the medium and low error rates was analyzed. Only one significant effect ($p \leq .10$) was found (at grade 5 mathematics). Furthermore, this effect was positive, but other effects that approached significance were negative. (See Technical Note V-1a, Marliave et al., 1977, for additional information on these analyses.) Therefore, it was concluded that the balance of low and medium error rates, represented by the product of the two, was unrelated to learning outcomes. This left the simple

conclusion that the medium error rate was negatively related with learning. This suggested that low error rate would be positively related to learning, because low and medium error are essentially ipsative. Therefore, analyses were conducted with low error rate.

Table 4-7 presents the results of analyses of low and high error rate in reading and mathematics at grades 2 and 5 during the A-B period. The proportion of residual variance accounted for by low and high error rate is reported for each content category. Note that unique proportions of residual variance are given, where shared variance accounted for by both engagement rate and low or high error rate is excluded from the computations. Similarly, where low error rate and high error rate accounted for the same residual variance, this was excluded from the unique proportions attributed to either error rate. However, the column reporting the unique proportion of variance for the low and high error rates together (the last entry to the right for each content category), includes the variance shared between the two error rates (but still excludes variance shared with engagement rate).

It should be noted, also, that whenever either one of the two error rate variables had a significant effect for a particular content outcome, then the joint variance accounted for by both variables together is also reported (even though the other of the two variables may have had a nonsignificant effect). In such cases, the sign of the nonsignificant effect is reported. However, the unique variance accounted for by the nonsignificant effect alone is not reported (because it is attributable to chance).

The results in Table 4-7 indicate that low error rate had a consistent positive relationship with learning outcomes, while high error rate was consistently negative. In no case was high error significantly negative or

Table 4-7

Regressions Analyzing the Effects of Low and High Error Rates

Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the A-B Period, Regressing Post Achievement on Pre Achievement, Engaged Time, Low Error Rate, and High Error Rate.

Grade 2 (N=127)									Grade 5 (N=122)								
Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Low and High Error Rate Effects ^a							Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Low and High Error Rate Effects ^a						
		* p ≤ .10 ** p ≤ .01		+ positive - negative		Unique Proportion of Residual Variance ^b					* p ≤ .10 ** p ≤ .01		+ positive - negative		Unique Proportion of Residual Variance ^b		
		Low	High	Low	High	Low	High	Low & High ^c			Low	High	Low	High	Low	High	Low & High ^c
Total Comprehension	.31		*	+	-		.04	.06	Total Word Meaning	.58	*		+	-	.05		.05
Total Decoding	.65								Total Comprehension	.71	**		+	+	.10		.10
Total Word Structure	.40	**		+	-	.10		.14	Word Structure Syllables	.41		*	+	-		.03	.03
Total Reading	.71	*		+	-	.02		.06	Total Reading	.81	**		+	+	.08		.08
Addition and Subtraction no Regrouping	.41	**		+	+	.08		.08	Total Geometry	.24	*	*	+	-	.03	.02	.06
Addition and Subtraction Speeded Tests	.49	*		+	+	.04		.05	Total Multiplication	.28	**	**	+	-	.06	.06	.14
Computational Transfer	.28	**	*	+	-	.04	.02	.12	Multiplication Speeded Tests, Basic Facts Time	.54							
Place Value and Numerals	.38								Division	.24							
Word Problems	.29								Fractions	.08	*		+	-	.02		.02
Money	.43								Computational Transfer	.55	*		+	-	.02		.05
Total Mathematics	.63	**		+	+	.08		.08	Word Problems	.22	*		+	-	.04		.07
									Total Mathematics	.70		**	+	-		.06	.08

^aBased on observation

^aBased on observation.

^bThis is the proportion of the residual variance accounted for uniquely by Low Error Rate or High Error Rate. Residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed."

^cThis is the unique proportion of residual variance accounted for by Low and High Error Rate, including variance "shared" among these two variables.

low error significantly positive. Furthermore, each of the two error rate variables had at least one significant effect in each of the four quadrants (reading and mathematics at grades 2 and 5). The two strongest cases are total word structure at grade 2 and total multiplication at grade 5, where low and high error rate jointly accounted for 14 percent of the residual variance.

Table 4-8 presents the results of analyses of engaged, low error rate time. Engaged, low error rate time is the product of engaged time and low error rate. (Recall that engaged time is itself the product of allocated time and engagement rate.) This variable was considered to be an estimate of the simultaneous occurrence of allocated time in a content category, engagement, and low error. Note that Table 4-8 uses the same format as Table 4-6.

The results of Table 4-8 indicate that engaged, low error rate shows consistent positive effects, as would be expected on the basis of the previous results for engaged time and low error rate. All significant effects are positive, except that for word problems at grade 2, which is one of the two smallest significant effects. Positive effects are reported for every quadrant except grade 5 reading.

However, the engaged, low error rate variable does not appear to be as effective as engaged time and low error rate entered separately. The sum of the residual variance accounted for by engaged time and low error rate (from Tables 4-6 and 4-7) is equivalent to their joint variance accounted for, including shared variance. (Recall that Table 4-6 gives the variance accounted for by engaged time without the entry of low error rate, while Table 4-7 gives the unique variance accounted for by low error after the variance for engaged time has been removed.) Summing the residual variance

Table 4-8

Regressions Analyzing the Effects of Engaged, Low Error Rate Time
Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the A-B Period, Regressing
Post Achievement on Pre Achievement and Engaged, Low Error Rate Time

Grade 2 (N=127)					Grade 5 (N=122)				
Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Engaged, Low Error Rate Time ^a Effects			Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Engaged, Low Error Rate Time ^a Effects		
		Significance * p ≤ .10 ** p ≤ .01	Sign + positive - negative	Proportion of Residual Variance ^b			Significance * p ≤ .10 ** p ≤ .01	Sign + positive - negative	Proportion of Residual Variance ^b
Total Comprehension	.31	**	+	.08	Total Word Meaning	.58			
Total Decoding	.65				Total Comprehension	.71			
Total Word Structure	.40	**	+	.07	Word Structure Syllables	.41			
Total Reading	.71	**	+	.08	Total Reading	.81			
Addition and Subtraction No Regrouping	.41	*	+	.04	Total Geometry	.24	**	+	.09
Addition and Subtraction Speeded Tests	.49				Total Multiplication	.28	**	+	.11
Computational Transfer	.28	**	+	.05	Multiplication Speeded Tests, Basic Facts Time	.54			
Place Value and Numerals	.38				Division	.24	*	+	.04
Word Problems	.29	*	-	.02	Fractions	.08	**	+	.30
Money	.43	*	+	.02	Computational Transfer	.55			
Total Mathematics	.63	**	+	.06	Word Problems	.22			
					Total Mathematics	.70			

^aComputed by taking the product of Allocated Time in the content category (based on teacher logs), Engagement Rate (based on observation), and Low Error Rate (based on observation).

^bThis is the proportion of the residual variance accounted for by Engaged, Low Error Rate Time, where residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed."

accounted for by engaged time and low error rate, an average of 6.5% of residual variance is accounted for across the 23 regressions. The average for engaged, low error rate time is 4.2%. Therefore, engaged time and low error rate were kept separate in subsequent analyses. It is interesting to note, however, that engaged, low error rate time accounts for the greater proportion of residual variance (30%) in the content category for which either approach achieves its maximum variance (fractions at grade 5). Engaged time and low error rate separately account for a total of 22% of the residual variance in this case.

B-C analyses. Table 4-9 presents the results of analyses of engaged time in reading and mathematics at grades 2 and 5 during the B-C period. This table follows the same format used in the preceding tables. Note that some changes were made in the set of test scales administered at grade 2 for the B-C period (as compared with those for the A-B period). For example, several scales were used for different decoding skills, as opposed to the single "total decoding" scale used for the A-B period. These changes were made to reflect changes in instructional emphases from one period to the next. As students progress in their decoding instruction at grade 2, they begin to accumulate enough knowledge of various decoding skills that tests differentiating these skills become useful. Similarly, grade 2 students in this study did not receive enough instruction on fractions during the A-B period to warrant its analysis. However, instructional time on fractions at grade 2 was sufficient during the B-C period to justify its inclusion in analyses for that period.

The B-C analyses of engaged time confirm the positive effects found with the A-B data. In addition, these B-C results are more consistent than those for the A-B period, in that positive effects were found in all four quadrants (reading and mathematics at grades 2 and 5, see Table 4-9). No significant

Table 4-9

Regressions Analyzing the Effects of Engaged Time
Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the B-C Period, Regressing
Post Achievement on Pre Achievement and Engaged Time

Grade 2 (N=139)					Grade 5 (N=122)				
Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Engaged Time ^a Effects			Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Engaged Time ^a Effects		
		Significance * $p \leq .10$ ** $p \leq .01$	Sign + positive - negative	Proportion of Residual ^b Variance			Significance * $p \leq .10$ ** $p \leq .01$	Sign + positive - negative	Proportion of Residual ^b Variance
Total Comprehension	.32				Total Word Meaning	.69			
Decoding Blends & Long Vowels	.49	*	+	.03	Total Comprehension	.66	**	+	.14
Decoding Variant Consonants	.12				Word Structure Syllables	.35			
Decod. Comp. Patt. Spelling Time	.47	**	+	.05	Total Reading	.77	**	+	.06
Word Structure Meaningful Units	.49				Total Geometry	.12			
Word Structure Syllables	.09	**	+	.07	Total Multiplication	.33			
Total Reading	.68	*	+	.02	Mult. Speed Test, Basic Facts Time	.65	**	+	.05
Add & Subt. No Regrouping	.24				Division	.30			
Add & Subt. Speeded Test	.35	*	+	.03	Fractions	.28	**	+	.26
Add & Subt. With Regrouping	.06	*	+	.03	Computational Transfer	.58			
Computational Transfer	.32				Word Problems	.38			
Place Value and Numerals	.47				Total Mathematics	.69	*	+	.03
Word Problems	.27	*	+	.03					
Money	.34								
Linear Measurement	.19								
Fractions	.05	**	+	.10					
Total Mathematics	.58								

^aComputed by taking the product of Allocated Time in the content category (based on teacher logs) and Engagement Rate (based on observation).

^bThis is the proportion of the residual variance accounted for by Engaged Time, where residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed."

negative effects were found for engaged time in the B-C period. Where the negative A-B result had occurred (multiplication speeded tests), a significant positive effect was found in the B-C analyses. The strongest effect was for fractions at grade 5, where 26 percent of the residual post achievement variance was accounted for by engaged time.

The second set of B-C regressions analyzed allocated time and engagement rate as two separate variables. As discussed above, it had been found in the A-B analyses that the separation of engaged time and low error rate was more effective than their combination in terms of predicting student learning. Therefore, it was thought that separating allocated time and engagement rate might also be useful.

Table 4-10 presents the results of analyses of allocated time and engagement rate in reading and mathematics at grades 2 and 5 during the B-C period. The format of this table is the same as that for Table 4-7. Consistent, positive results were obtained for both allocated time and engagement rate. Significant positive effects were found for both variables in all quadrants. In no case did a significant negative effect occur. In the strongest case, fractions at grade 5, allocated time and engagement jointly accounted for 26% of the residual variance (the same percentage found for engaged time). Table 4-10 shows, overall, that allocated time and engagement rate, as two separate variables, are more highly predictive of student learning than their product, engaged time. Comparing this table with table 4-9, it is apparent that the two separate variables taken jointly often account for more variance than their single product, and they never account for less. The two separate variables jointly account for an average of 4.9% of the residual variance, while their single product accounts for an average of 3.1%. Therefore, it was concluded that allocated time and engagement

Table 4-10

Regressions Analyzing the Effects of Allocated Time and Engagement Rate
Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the B-C Period, Regressing
Post Achievement on Pre Achievement, Allocated Time, and Engagement Rate

Grade 2 (N=139)									Grade 5 (N=122)								
Content Category for Post Achievement, Pre Achievement, and Allocated Time	Proportion of Total Variance Accounted for by Pre Achievement	Allocated Time ^a and Eng. Rate ^b Effects							Content Category for Post Achievement, Pre Achievement, and Allocated Time	Proportion of Total Variance Accounted for by Pre Achievement	Allocated Time ^a and Eng. Rate ^b Effects						
		* p ≤ .10 ** p ≤ .01		+ positive - negative		Unique Proportion of Residual Variance ^c					* p ≤ .10 ** p ≤ .01		+ positive - negative		Unique Proportion of Residual Variance ^c		
		Time	Eng.	Time	Eng.	Time	Eng.	Time & Eng. ^d			Time	Eng.	Time	Eng.	Time	Eng.	Time & Eng. ^d
Total Comprehension	.32								Total Word Meaning	.69		**	-	+		.15	.15
Decoding Blends and Long Vowels	.49		**	+	+		.08	.09	Total Comprehension	.66	**		+	-	.15		.15
Decoding Variant Consonants	.12		*	+	+		.02	.03	Word Structure Syllables	.35							
Decod. Comp. Patt. Spelling Time	.47	*	*	+	+	.03	.03	.07	Total Reading	.77	*	*	+	+	.02	.02	.06
Word Structure Meaningful Units	.49								Total Geometry	.12		*	+	+		.03	.04
Word Structure Syllables	.09	**		+	+	.05		.08	Total Multiplication	.33							
Total Reading	.68		*	+	+		.03	.04	Mult. Speed Test, Basic Facts Time	.65	**	*	+	+	.06	.04	.10
Add & Subt. No Regrouping	.24								Division	.30							
Add & Subt. Speeded Test	.35	*	**	+	+	.02	.04	.07	Fractions	.28	**		+	+	.24		.26
Add & Subt. With Regrouping	.06	*	*	+	+	.03	.02	.04	Computational Transfer	.58							
Computational Transfer	.32								Word Problems	.38							
Place Value and Numerals	.47								Total Mathematics	.69	**		+	+	.06		.06
Word Problems	.27	*		+	+	.02		.04									
Money	.34		*	+	+		.04	.04									
Linear Measurement	.19																
Fractions	.05	**		+	-	.10		.10									
Total Mathematics	.38																

^aBased on teacher logs.

^bBased on observation.

^cThis is the proportion of residual variance accounted for uniquely by Allocated Time or Engagement Rate. Residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement has been "removed."

^dThis is the residual variance accounted for by Allocated Time and Engagement Rate, including variance "shared" among these two variables.

rate are more strongly related to student learning when analyzed as two separate variables. Subsequent analyses entered both allocated time and engagement rate.

The A-B analyses of low and high error rate were replicated on the B-C data with one difference. Error rate for the A-B analyses had been aggregated over all reading and all mathematics, including developmental and related content for which no achievement scales were administered. Engagement rates had been computed only for the reading and mathematics content categories with some direct relation to the achievement scales administered. Similarly, total reading and mathematics allocated time were computed to exclude content not directly related to the tests used. It was decided that this procedure should be followed with low and high error rate analyses of the B-C data. As a preliminary examination of the differences between the two procedures for computing error rate, correlations were obtained for low and high error computed by each procedure. These correlations ranged from .93 to .98. In addition, the regressions analyzing the relationship of error rate to student learning were conducted on low and high error rates computed by both procedures. As suggested by the previous high correlations, the results of these analyses were substantially the same for both procedures. (See Technical Note V-2a, Marliave, Fisher, & Dishaw, 1978, for further information on these comparisons.) Nevertheless, the aggregation of error rates for just those content categories that correspond to the achievement scales administered, appeared to be logically more appropriate. Therefore, the results presented below pertain to those error rates.

Table 4-11 presents the results of analyses of low and high error rate during the B-C period. This table follows the same format as Table 4-7.

Table 4-11

Regressions Analyzing the Effects of Low and High Error Rates

Analyses of Reading and Mathematics Achievement at Grades 2 and 5 During the B-C Period, Regressing Post Achievement on Pre Achievement, Engaged Time, Low Error Rate, and High Error Rate

Grade 2 (N=139)									
Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Low and High Error Rate Effects ^a							
		* p ≤ .10 ** p ≤ .01		+ positive - negative		Unique Proportion of Residual Variance ^b			
		Low	High	Low	High	Low	High	Low & High ^c	
Total Comprehension	.32								
Decoding Blends and Long Vowels	.49	*	**	+	-	.03	.06	.12	
Decoding Variant Consonants	.12	**		+	-	.08		.11	
Decoding Complex Patterns, Spelling Time	.47	**		+	-	.10		.15	
Word Structure Meaningful Units	.49								
Word Structure Syllables	.09	*		+	-	.04		.07	
Total Reading	.68	*	*	+	-	.03	.02	.08	
Addition and Subtraction No Regrouping	.24	*		+	-	.03		.04	
Addition and Subtraction Speeded Test	.35	**		+	-	.15		.19	
Addition and Subtraction With Regrouping	.06	*		+	+	.02		.03	
Computational Transfer	.32	**		+	+	.05		.06	
Place Value and Numerals	.47								
Word Problems	.27		*	+	-		.02	.04	
Money	.34	*	**	+	-	.02	.04	.09	
Linear Measurement	.19								
Fractions	.05								
Total Mathematics	.58	*		+	-	.04		.05	

Grade 5 (N=122)									
Content Category for Post Achievement, Pre Achievement, and Engaged Time	Proportion of Total Variance Accounted for by Pre Achievement	Low and High Error Rate Effects ^a							
		* p ≤ .10 ** p ≤ .01		+ positive - negative		Unique Proportion of Residual Variance ^b			
		Low	High	Low	High	Low	High	Low High ^c	
Total Word Meaning	.69		*	0	-		.02	.02	
Total Comprehension	.66		**	-	-		.05	.06	
Word Structure Syllables	.35		*	-	-		.04	.05	
Total Reading	.77		*	-	-		.03	.06	
Total Geometry	.12		*	-	-		.03	.03	
Total Multiplication	.33		**	+	-		.06	.09	
Multiplication Speeded Test, Basic Facts Time	.65	*		+	+	.04		.04	
Division	.30		**	-	-		.06	.06	
Fractions	.28		**	-	-		.04	.05	
Computational Transfer	.58	*	**	+	-	.02	.05	.09	
Word Problems	.38								
Total Mathematics	.69		*	+	-		.02	.02	>.0

^aBased on observation.^bThis is the proportion of the residual variance accounted for uniquely by Low Error Rate or High Error Rate. Residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement has been "removed."^cThis is the unique proportion of residual variance accounted for by Low and High Error Rate, including variance "shared" among these variables.

Significant positive effects were found for low error rate in all quadrants except grade 5 reading. Negative effects were found for high error rate in all four quadrants. No significant negative effect for low error or significant positive effect for high error was found. In the strongest case, addition and subtraction speeded tests at grade 2, 19% of the residual variance in post achievement was accounted for by low and high error rate jointly. Overall, then, the results of the A-B error rate analyses were replicated.

The final set of analyses on ALT and learning was a series of regressions examining the joint effect of all four ALT variables together (allocated time, engagement rate, low error rate, and high error rate). These analyses were conducted for both the A-B and the B-C period data sets, so that findings based on this final regression model could be compared across the two data sets. The results for the A-B period are presented in Table 4-12, while those for the B-C period are in Table 4-13. These tables follow, for the most part, the same format as the preceding tables. Where a variable showed a significant effect ($p \leq .10$), the unique proportion of residual variance accounted for is reported. In addition, the sign of all effects, regardless of significance, is reported. Finally, the joint contribution of all four ALT variables is reported for each content category (even when no variable shows a unique individual effect that is significant).

Summaries of the results in Tables 4-12 and 4-13 are presented in Tables 4-14 and 4-15. These simplified versions present results by quadrant, including the total number of regressions in the quadrant, the number (and sign) of significant effects for each of the four ALT variables, and the average residual variance accounted for by the four variables jointly.

The results presented in Tables 4-12 through 4-15 indicate that allocated time, engagement rate, and low error rate are each positively associated with

Table 4-12

Regressions Analyzing the Combined Effects of ALT Variables

Analyses of Reading and Mathematics Achievement in Grades 2 and 5 for the A-B Period, Regressing Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate

GRADE 2 (N=127)						GRADE 5 (N=122)					
Content Category for Post Achievement, Pre Achievement, and Allocated Time	Proportion of Variance Accounted for by Pre Achievement	Unique Proportion of Residual Variance ^a				Content Category for Post Achievement, Pre Achievement, and Allocated Time	Proportion of Variance Accounted for by Pre Achievement	Unique Proportion of Residual Variance ^a			
		+positive effect Alloc. Time ^b	+positive effect Eng. Rate ^c	+positive effect Low Error ^c	+positive effect High Error ^c			+positive effect Alloc. Time ^b	+positive effect Eng. Rate ^c	+positive effect Low Error ^c	+positive effect High Error ^c
Total Comprehension	.31	+	+	+	-	Total Word Meaning	.58	-	+	+	+
Total Decoding	.65	-	+	+	+	Total Comprehension	.7*	-	+	+	+
Total Word Structure	.40	+	+	+	+	Word Structure Syllables	.41	-	+	+	-
Total Reading	.71	+	+	+	+	Total Reading	.81	-	+	+	+
Addition and Subtraction No Regrouping	.41	+	+	+	+	Total Geometry	.24	+	+	+	-
Addition and Subtraction Speeded Tests	.49	-	+	+	+	Total Multiplication	.28	+	+	+	+
Computational Transfer	.28	+	+	+	+	Multiplication Speeded Tests, Basic Facts Time	.54	-	+	-	-
Place Value and Numerals	.38	+	+	+	+	Division	.24	+	+	+	-
Word Problems	.29	-	+	+	+	Fractions	.08	+	+	+	-
Money	.43	+	+	+	-	Computational Transfer	.55	-	+	+	+
Total Mathematics	.63	+	+	+	+	Word Problems	.22	+	+	+	+
						Total Mathematics	.70	-	+	+	+

^aThis is the proportion of the residual variance accounted for uniquely by the independent variable specified, where residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed."

^bBased on teacher logs.

^cBased on observation.

^dThis is the residual variance accounted for by Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate, including variance "shared" among these four variables.

Table 4-13

Regressions Analyzing the Combined Effects of ALT Variables

Analyses of Reading and Mathematics Achievement in Grades 2 and 5 for the B-C Period, Regressing Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate

Grade 2 (N=139)							Grade 5 (N=122)						
Content Category for Post Achievement, Pre Achievement, and Allocated Time	Proportion of Total Variance Accounted for by Pre Achievement	Unique Proportion of Residual Variance ^a				Residual Variance, Combined ALT Variables ^d	Content Category for Post Achievement, Pre Achievement, and Allocated Time	Proportion of Total Variance Accounted for by Pre Achievement	Unique Proportion of Residual Variance ^a				Residual Variance, Combined ALT Variables ^d
		+ positive effect		- negative effect					+ positive effect		- negative effect		
		Alloc. Time ^b	Eng. Rate ^c	Low Error ^c	High Error ^c				Alloc. Time ^b	Eng. Rate ^c	Low Error ^c	High Error ^c	
Total Comprehension	.32	+	+	+	-	.03	Total Word Meaning	.69	-	+	+	-	.15
Decoding Blends & Long Vowels	.49	+	+	+	-	.17	Total Comprehension	.66	+	+	-	-	.21
Decoding Variant Consonants	.12	+	+	+	-	.13	Word Structure Syllables	.35	-	+	-	-	.05
Decoding Complex Patterns, Spelling Time	.47	+	+	+	-	.21	Total Reading	.77	+	+	-	-	.11
Word Structure Meaningful Units	.19	+	+	+	-	.03	Total Geometry	.12	+	+	-	-	.07
Word Structure Syllables	.09	+	+	+	-	.14	Total Multiplication	.33	+	-	+	-	.11
Total Reading	.68	+	+	+	-	.11	Mult. Speed Test Basic Facts Time	.65	+	+	+	+	.13
Add & Subt. No Regrouping	.24	-	+	+	-	.04	Division	.30	+	-	-	-	.09
Add & Subt. Speeded Test	.35	+	+	+	-	.22	Fractions	.28	+	+	-	-	.30
Add & Subt. With Regrouping	.06	+	+	+	+	.06	Computational Transfer	.58	-	-	+	-	.10
Computational Transfer	.32	+	+	+	+	.06	Word Problems	.38	-	0	+	-	.01
Place Value and Numerals	.47	-	-	+	-	.01	Total Mathematics	.69	+	+	+	-	.08
Word Problems	.27	+	+	+	-	.07							
Money	.34	-	+	+	-	.12							
Linear Measurement	.19	0	+	+	-	.04							
Fractions	.05	+	+	+	-	.11							
Total Mathematics	.58	+	+	+	-	.06							

^a This is the proportion of the residual variance accounted for uniquely by the independent variable specified, where residual variance refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed."

^b Based on teacher logs.

^c Based on observation.

^d This is the residual variance accounted for by Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate, including variance "shared" among these four variables.

Table 4-14

Summary of Regressions
Combined Effects of ALT Variables

Regression of Post Achievement on Pre Achievement, Allocated Time,
Engagement Rate, Low Error Rate, and High Error Rate (A-B Period).

Grade	Content Domain	Total Number of Regressions	Number of significant effects ($p \leq .10$) positive + or negative -				Average Residual Variance, Combined ALT Variables ^a
			Allocated Time	Engagement Rate	Low Error Rate	High Error Rate	
2 (N=127)	Reading	4	2+	3+	2+	0	.07
2 (N=127)	Mathematics	7	1+/1-	1+	4+	1-	.04
5 (N=122)	Reading	4	0	1+	3+	0	.03
5 (N=122)	Mathematics	8	4+	2+	3+	4-	.09

^a"Residual variance" refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed." This column reports the proportion of this residual variance that is accounted for by the set of ALT variables.

Table 4-15

Summary of Regressions

Combined Effects of ALT Variables

Regression of Post Achievement on Pre Achievement, Allocated Time, Engagement Rate, Low Error Rate, and High Error Rate (B-C Period).

Grade	Content Domain	Total Number of Regressions	Number of significant effects ($p \leq .10$) positive + or negative -				Average Residual Variance, Combined ALT Variables ^a
			Allocated Time	Engagement Rate	Low Error Rate	High Error Rate	
2 (N=139)	Reading	7	3+	2+	5+	2-	.12
2 (N=139)	Mathematics	10	4+	0	5+	2-	.08
5 (N=122)	Reading	4	1+	1+	0	3-	.13
5 (N=122)	Mathematics	8	3+	2+	2+	6-	.11

4-33

^a "Residual variance" refers to the variance remaining in the Post Achievement measure after Pre Achievement is "removed." This column reports the proportion of this residual variance that is accounted for by the set of ALT variables.

achievement outcomes. High error rate is negatively associated with achievement. These findings were consistent for analyses of both the A-B and B-C data sets, although the B-C period findings were slightly stronger and more consistent than those for the A-B period. Of the significant effects ($p \leq .10$), there was only one exception to generalization that effects for allocated time, engagement, and low error rate were positive, while those for high error rate were negative. This exception was the negative effect for allocated time in addition and subtraction speed tests at grade 2 during the A-B period. Note, however, that allocated time showed a positive effect in the same content area during the B-C period. Further, this positive B-C effect accounted uniquely for 3% of the residual variance in achievement, while the negative A-B effect accounted for only one percent of the residual variance.

The slightly greater consistency and strength of the B-C results (as compared with the A-B results) is evident upon examination of the summary tables (Tables 4-14 and 4-15). As noted above, the exceptional negative effect for allocated time occurred in the A-B results. In addition, there were more cases in the A-B results where an ALT variable showed no significant effects in a particular quadrant. Allocated time in grade 5 reading and high error rate in both grade 2 reading and grade 5 reading showed no significant effects during the A-B period. No significant effects during the B-C period were found for engagement rate in grade 2 mathematics and low error rate in grade 5 reading. (It is interesting to note that each of the four ALT variables showed significant effects in all four quadrants during one of the two periods, but not the other.)

The greater strength of the B-C results is shown by the average residual variance of the combined ALT variables for each of the quadrants (the

right-hand columns of Tables 4-14 and 4-15). These averages, for A-B and B-C findings respectively, were .07 and .12 in grade 2 reading, .04 and .08 in grade 2 mathematics, .03 and .13 in grade 5 reading, and .09 and .11 in grade 5 mathematics. In each case, the B-C results are stronger (in terms of average residual variance accounted for by the combined ALT variables).

Overall, the results are quite consistent in support of the positive effects for allocated time, engagement rate, and low error rate, as well as the negative effect for high error rate. A total of 52 regressions were conducted. Of the 73 significant effects found, only one was contrary to the general trend of the four ALT variables. Furthermore, in the B-C analyses where the results were stronger, the set of ALT variables accounted for an average of about 11% of the residual variance in achievement, while more than 20% of the residual variance was accounted for with one regression in each quadrant. In the strongest example, fractions at grade 5 (B-C period), the ALT variables accounted for 30% of the residual variance.

Discussion: ALT and Student Learning

Data have been presented showing that allocated time, engagement rate, and low error rate are positively associated with residual achievement, while high error rate is negatively associated with residual achievement. Several aspects of these findings warrant discussion at this point.

Non-experimental. First, it should be recognized that these data are non-experimental. No random assignment or systematic manipulation of treatments was undertaken. The reader should always bear in mind that the findings reported here cannot show causal relationships directly. Rather, these findings indicate that one or more variables are associated. The sign of that association (positive or negative) and its relative magnitude are also represented in these analyses.

However, it is worth recognizing that one may reasonably base a belief on inconclusive evidence. No knowledge is ever absolute. Even experimental analyses are generally open to more than one reasonable interpretation, particularly when one wishes to generalize to natural situations and events. Correlative data combined with experiential knowledge and logical reasoning often provide considerable evidence for causal relationships. One should recognize the limitations of such evidence, but not disregard it.

Strength of the findings. Another aspect of these findings that should be discussed is their strength. The effects of the ALT variables were not surprisingly large. Generally, the four ALT variables combined accounted for about a tenth of the residual variance in achievement. However, these effects were unusually consistent. As noted previously, only one of 73 significant effects was contrary to the general trends of these variables.

It is also encouraging that none of these general trends was counter-intuitive. Experience and logic would support the finding that allocated instructional time was positively associated with student learning. Similarly, one would expect that higher levels of student engagement during instruction should be positively related to learning, while the student's complete misunderstanding of the instruction (high error rate) should be negatively related to learning. The positive association of low error rate and learning is in accordance with the tenets of programmed learning theory. However, this finding was not predicted and is the one association of ALT with learning that is not clearly substantiated by experience and logic. For this reason, additional consideration of the findings for low error rate will be presented later in this discussion.

It should also be noted that the true strength of the effects for ALT is probably greater than the tenth of residual achievement variance shown for the data presented here. This is due to the fact that measurement errors create random variance that nothing could account for systematically. If one could eliminate this erroneous variance, then a greater proportion of the remaining variance would be accounted for by the four ALT variables.

It is possible to make a rough estimate of what the proportion of variance accounted for would be if measurement error could be eliminated. As an illustration, consider the variance in residual post achievement. One can estimate the reliability of this residual variable,¹ which gives a rough estimate of the proportion of its variance that is true variance (the remaining proportion being error variance). If this reliability is .50, then half of the variance in residual achievement is estimated to be true, while the other half is estimated to be error. As mentioned above, however, nothing could account for the half that is error variance. Therefore, if a variable accounted for 10% of the total residual variance, then it was actually accounting for 10% out of a possible 50%. Therefore, 20%, i.e., $10 \div 50$, is the estimated proportion of the true residual variance that was accounted for by this variable.

As an example of the importance of measurement error, consider the residual variance accounted for by the combined ALT variables in syllables at grade 2 during the B-C period. This was 14% (Table 4-13). The reliability of residual syllables achievement (grade 2, B-C period) was estimated at .44. This leads to the estimation that 32% of the residual variance would have been accounted for if error of measurement could have been eliminated from the syllables scores. For fractions at grade 5 during the B-C period, the combined ALT variables accounted for 30% of the residual

¹The formula for the reliability of a residual is given by O'Connor (1972, page 92, formula 50).

variance (Table 4-13). Residual fractions achievement (grade 5, B-C period) had an estimated reliability of .68. This results in the estimation that 44% of the residual variance would have been accounted for if error of measurement were eliminated from the residual scores. Note that ALT on fractions accounted for more than twice as much residual variance as ALT on syllables in the analyses conducted. However, it is estimated that ALT on fractions would account for less than one and one-half times as much residual variance as ALT on syllables if error of measurement could be eliminated.

One final point regarding the strength of effects for the ALT variables is that one tenth of student learning over a school year can be an effect of considerable importance. This is especially true when such an effect can be attributed to a variable that is under the control of the classroom teacher. As an example, consider the mathematics achievement of grade 5 students in this study over the 15-week B-C period. The combined ALT variables accounted for 8% of the residual mathematics achievement variance at grade 5 during the B-C period. Allocated time alone accounted for 6% of this residual variance. The average score for the total mathematics test at the B testing occasion was 70, while that at the C testing occasion was 88. Hence, the average student gained 18 items in total mathematics at grade 5 over the B-C period. This was accomplished with an average of 32 minutes per school day of instructional time allocated to mathematics directly relevant to the test. The average engagement during this instruction was 73%. The average student spent 34% of this instructional time at the low error response category and 3% of this time at the high error category. Using the regression coefficients generated by the analysis of these data (Marliave, Fisher, and Dishaw, 1978), it is possible to estimate how much the C test score of this "average" student could be increased by

increasing the amount of allocated instructional time in mathematics. If this allocated time were increased from the average of 32 minutes per day to 64 minutes per day, the estimated C test score would be 103. This is an increase of 15 items (about half of a standard deviation).

It is also possible to apply this estimation procedure to an atypical student. The lowest score that occurred on the same test (total mathematics at grade 5) at the B testing occasion was 2.7 (almost zero). The estimation procedure would predict a C test score of 31 if this student received the average amount of allocated mathematics instruction, had an average engagement rate, an average low error rate, and an average high error rate. Suppose a teacher decided to allocate massive quantities of mathematics instruction for this student in an attempt to remediate his/her deficient achievement. The estimation procedure would predict that by allocating about $2\frac{1}{2}$ hours of mathematics instruction per day, rather than the average of $\frac{1}{2}$ hour, this student would attain about the average mathematics achievement score on the C test (85). Two and one-half hours per day is a lot of mathematics instruction, but improving from lowest to average achievement in 15 weeks is a lot of learning.

These estimations of the effects of hypothetical changes in allocated instructional time are purely speculative. The procedure is entirely theoretical, generating linear extrapolations from one sample of data. Furthermore, causal relationships are assumed to underlie associative effects. However, these examples do provide concrete illustrations of the possible effects of changes in allocated instructional time. They indicate that intentional manipulations of one ALT variable, allocated time, could produce substantial and important changes in student achievement.

Generalizability. As discussed previously, attribution of causal relationships to variables correlated in this study requires inference and

must be treated as tentative. The preceding discussion is based on the inference that the association of allocated time and student learning is attributable to the causation of learning by allocated time. However, even if one accepts this inference as reasonable, despite the limitations of correlative evidence, it should also be recognized that there are limitations to the generalizability of such relationships. For example, even if it is true that allocated time causes learning, it only does so under certain conditions and up to some limit. If the student engagement rate is zero, then no amount of allocated time will cause any learning. Furthermore, even under the best of conditions, there is a limit to the amount of work that a student can accomplish within a day. Therefore, the allocation of twenty hours per day of mathematics instruction might produce less learning than a more moderate dose of instruction.

It is particularly important that the findings for error rate not be extrapolated to extremes. The average student in this study spent no more than half of his/her time working at a low error rate level. Relative to this average, more time at the low error rate level was associated with greater increases in achievement. However, it is quite possible that very high proportions of low error rate instruction (such as 80 or 90%) would be negatively associated with learning outcomes. A moderate proportion of instruction at the medium error rate level, where the student is exposed to some problems that he/she is unable to solve, may be advantageous. The issue is to identify the optimal mix of tasks in terms of student error rates. This study indicates only what was associated with learning among the varieties of instruction sampled. The relationships for error rate found here may change when one extrapolates to situations that rarely or never occurred in the classrooms sampled for this study.

One should also be careful not to assume that limitations apply in areas where there is no evidence that this is so. For example, it may appear reasonable to assume that allocated time will be more highly related to learning outcomes for more narrowly delimited skill areas, such as division, than broadly conceived abilities, such as reading comprehension. However, the second largest effect for allocated time was in total reading comprehension at grade 5 during the B-C period (see Table 4-13), where 14% of the residual variance was accounted for by time. Time allocated to division had no significant effect during the B-C period, although it did account for 2% of the residual variance during the A-B period (Tables 4-12 and 4-13). This is not to say that the more broadly defined content areas showed greater effects for allocated time. The largest effect for allocated time was for fractions at grade 5 during the A-B period (22% of residual variance).

One might also be tempted to assume that stronger effects would appear at grade 2 than at grade 5, or in mathematics rather than in reading. Generally, such trends did not appear consistently in the findings. For example, allocated time in mathematics did appear to account for greater proportions of variance than reading time, but only at grade 5. Furthermore, even at grade 5, the combined effects for ALT in reading were comparable to those in mathematics (see Tables 4-12 and 4-13).

Follow-up analyses of low error rate. The findings for low error rate are probably the most difficult to interpret among the relationships between ALT and achievement. Common sense would readily suggest that allocated time and engagement rate should be positively associated with student learning outcomes, while high error rate should be negatively associated with learning. Furthermore, a relatively simple case can be made for the notion that increases in allocated time and engagement should cause higher levels

of student learning, except when high error rates on instructional activities inhibit learning. Correlative analyses, such as those presented here, can never demonstrate causal relationships. However, it is reasonable to infer that the associations of residual achievement with allocated time, engagement rate, and high error rate show that learning is facilitated by student time spent paying attention to instructional tasks, where the student is not totally confused by the tasks.

The findings for low error rate are more difficult to interpret because either of two opposite causal relationships may be inferred from the correlative results.² It was found that more instruction at a low student error rate tended to be associated with higher levels of student learning (residual achievement). One reasonable interpretation is that students learn more as a result of spending relatively greater amounts of time on tasks where they have a low error rate. However, it is also plausible that students who had low error rates were generally those with the highest skills and abilities, so that the higher levels of learning attained by these students were due to these relatively high skills and abilities.

To some extent, the second interpretation is contradicted by the fact that the effects of error rate were computed after accounting for pre achievement. Pre achievement should be one of the best measures of entering student skills and abilities for learning the content of an achievement scale. However, no measure is perfect. Even after pre achievement has been accounted for, there may be remaining effects due to skills and abilities that the pre achievement scale did not measure. Low error rate could reflect some of these unmeasured skills and abilities. It is possible, then, that the

²The complexity of the positive low error rate effect is underscored by the fact that BTES staff originally expected either negative or mixed effects for low error rate.

positive effect of low error rate on residual achievement was due to its association with skills and abilities that were not measured by pre achievement.

There is no means by which one can determine with certainty what explanation underlies the positive association of low error rate with student learning. However, the data leave some clues indicating whether the positive effects of low error rate really reflect instructional characteristics, rather than simply student skills and abilities.

One clue is the relationship between: 1. The proportion of total variance in post achievement that is accounted for by pre achievement, and 2. the proportion of residual variance in post achievement that is accounted for uniquely by low error rate. Pre achievement scores reflect entering student skills and abilities to the extent that these were measured. If the positive effects of low error rate were due to its reflection of skills and abilities that were not measured by the pre achievement scale, then low error rate should account for more variance when pre achievement accounts for less variance. That is, when pre achievement accounts for less variance, then there is probably more variance remaining that a second measure of student skills and abilities could reflect.

The relationship between the variance accounted for by pre achievement and that accounted for by low error rate can be examined in Tables 4-12 and 4-13. As discussed above, we would expect low error rate to account for larger proportions of variance in those content categories where the pre achievement scale accounted for less variance, if low error rate represents student skills and abilities rather than characteristics of instruction. There is one case in Table 4-12 where pre achievement accounted for less than 10% of the post achievement variance. This was fractions at grade 5.

In this case, the fractions achievement scale was not administered as a pre test (at A). Therefore, another scale (computational transfer) had to be used to represent pre achievement in fractions. This scale accounted for little post achievement variance (8%). Nevertheless, low error rate had no significant effect for fractions at grade 5 in the A-B period. The largest effect due to low error rate during the A-B period (Table 4-12) was for total word structure at grade 2, where 8% of the variance was accounted for uniquely by low error rate. However, in this case the pre achievement scale accounted for 40% of the post achievement variance. Furthermore, low error rate had a significant effect in the A-B period regression where pre achievement had its largest effect. This was total reading at grade 5, where pre achievement accounted for 81% of the post achievement variance, while low error rate accounted uniquely for 2% of the residual variance.

Table 4-13 also shows that low error rate did not have effects only where pre achievement was a weak predictor. There are three cases shown in this table where pre achievement accounted for less than 10% of the post achievement variance. These are syllables, addition and subtraction with regrouping, and fractions, all at grade 2. The pre and post achievement scales were identical for syllables, partially different for addition and subtraction with regrouping, and completely different for fractions. Low error rate had a significant effect only in the case of syllables (accounting for 4% of the residual variance). The strongest effect for low error rate during the B-C period was for addition and subtraction speeded tests at grade 2, where 14% of residual variance was accounted for by low error. Pre achievement accounted for 35% of post achievement variance in this case.

As a rough estimate of the overall relationship between the variance accounted for by pre achievement and that accounted for by low error rate, these two proportions can be correlated across the 52 regressions presented

in Tables 4-12 and 4-13. This yields a correlation of $-.06$, which is nonsignificant and show essentially no relationship. This indicates that the magnitude of effect of low error rate was largely independent of the accuracy with which the pre achievement scale predicted post achievement. This, in turn, supports the conclusion that low error rate is not simply a second measure of entering student skills and abilities.

A second clue to the reasons for the positive association of low error rate with student learning may be derived from the data displayed in Table 4-16. This table shows the conjoint relationships, in terms of residual achievement over the A-C period, of pre achievement at the A testing and low error rate over the A-C period. Data are shown for each of the four quadrants separately. Rows represent three levels of pre achievement (high, medium, and low pre test scores), while columns represent three levels of low error rate (high, medium, and low proportions of time spent at the low error rate). Cell entries show the mean residual achievement for students in that cell, as well as the number of students included in the cell.

Several relationships are made evident by Table 4-16. First, it is interesting to note that at grade 2 mathematics and grade 5 reading, students at the medium level of pre achievement had a higher average residual achievement than those at the high or low pre achievement levels. The average residual achievement in each quadrant must, as a mathematical artifact, be zero. The positive residuals for students at the medium levels of pre achievement in grade 2 mathematics and grade 5 reading indicate that there was a curvilinear relationship between pre and post achievement in these quadrants.

Table 4-16 also shows that pre achievement and low error rate were positively associated, because there were more students in the cells for which both pre achievement and low error rate are high, medium, and low than there were

Table 4-16

Conjoint Relationship of Pre-Achievement (at the A Testing)
and Low Error Rate (A-C Period) in Terms of
Residual Achievement (C Achievement Residualized on A Achievement).

Grade 2 (N=139)

		Reading S.D. = 21.8 ^a					Mathematics S.D. = 23.5 ^a		
Pre-Achievement (A Testing)	High	-13 (n=3)	4 (n=22)	0 (n=22)		High	-11 (n=12)	-10 (n=14)	3 (n=21)
	Medium	-6 (n=13)	-3 (n=19)	5 (n=12)		Medium	4 (n=10)	4 (n=18)	12 (n=15)
	Low	-10 (n=23)	3 (n=19)	37 (n=6)		Low	-7 (n=24)	-8 (n=12)	15 (n=13)
		Low	Medium	High			Low	Medium	High
Low Error Rate ^b					Low Error Rate ^b				

Grade 5 (N=122)

		Reading S.D. = 25.8 ^a					Mathematics S.D. = 27.8 ^a		
Pre-Achievement (A Testing)	High	2 (n=7)	-3 (n=27)	-1 (n=16)		High	0 (n=10)	-2 (n=16)	2 (n=13)
	Medium	2 (n=13)	2 (n=24)	11 (n=6)		Medium	1 (n=13)	-2 (n=13)	2 (n=14)
	Low	-5 (n=15)	-2 (n=25)	3 (n=6)		Low	1 (n=13)	-3 (n=24)	8 (n=11)
		Low	Medium	High			Low	Medium	High
Low Error Rate ^b					Low Error Rate ^b				

Cell entries are mean residual achievement for the A-C period and the number of students in the cell ("n" in parentheses).

^a"S.D." refers to the standard deviation of residual achievement, which is given for each quadrant.

^bFor all quadrants except grade 5 mathematics, less than 40 percent low error rate was categorized as a low percentage (left column), while greater than or equal to 60 percent low error rate was categorized as a high percentage (right column). In grade 5 mathematics, these categories were less than 20 percent and greater than 45 percent, respectively.

students in the cells for high pre achievement and low error rate or high error rate and low pre achievement. This indicates that low error rate, in part, reflects student pre achievement. Students who begin the year knowing more make fewer errors on daily tasks. However, low error rate may also be related to characteristics of instruction, such as the extent to which tasks are matched to the skills and abilities of students. The residual achievement data in Table 4-16 help clarify that part of low error rate that is not related to pre achievement.

If one examines the average residual achievement for each of the three columns in each quadrant, it is apparent that the residual achievement becomes greater at higher levels of low error rate. This represents the positive association of low error rate with learning. Next, an examination of the average residual achievement for each of the three rows in each quadrant indicates that there is no consistent relationship between pre achievement and residual post achievement. As noted above, there is a non-linear relationship between pre achievement and residual achievement at grade 2 mathematics and grade 5 reading. However, this does not appear in the other two quadrants. Furthermore, as a mathematical artifact of the regression procedures, pre achievement cannot have a linear relationship with residual post achievement (residualized on pre achievement).

If the positive association of low error rate and residual achievement were due to low error rate's association with student skills and abilities that were not measured by the pre-achievement test, then residual achievement would be approximately zero in those cells for which pre achievement and low error rate agreed. That is, where pre achievement and low error rate were both low, both medium, or both high, then both measures would be indicating the same relative level of skills and abilities for students. This would

indicate that there is relatively little student skill and ability associated with low error rate that is not measured by pre achievement, in those cells. Both measures may fail to detect certain differences in student skills and abilities, but where the two measures agreed, there would not be skill and ability measured by one but not the other.

However, if low error rate represented an instructional treatment, rather than just a second skill and ability measure, then it would make no sense to even think of "agreement" between the pre achievement and low error rate scores. The two variables would be measuring completely different factors, one the entering skills and abilities of students and the other an aspect of the instructional treatment.

The results of Table 4-16 show that where pre achievement and low error rate were both low, both medium, or both high, agreement between the two measures is not represented. In all quadrants except grade 5 mathematics, there is a strong trend toward negative residual post achievement where pre achievement and low error rate are both low.

It is possible that these results show an aptitude-by-treatment-like interaction, where the effects of the treatment variable (low error rate) are greater for students who are lower in pre achievement. This is apparent when one examines the difference in residual achievement between the low and the high level of low error rate, within a given level of pre achievement (i.e., within a given row). For example, consider grade 2 reading. Within the row for high pre achievement, the average residual post achievement for the low level of low error rate is -13. Within the same row, the average for the high level of low error rate is 0. The difference is 13. The corresponding difference for the medium level of pre achievement is 11. That for the low level of pre achievement is 47. Clearly, the difference in residual

achievement between the low and high level of low error rate is dramatically greater at the low level of pre achievement than at the medium or high levels of pre achievement. This trend is apparent in each of the other three quadrants. The corresponding differences for high, medium, and low levels of pre achievement, respectively, are 14, 8, and 22 at grade 2 mathematics. Those at grade 5 reading are 3, 8, and 8, while those at grade 5 mathematics are 2, 2, and 7. Hence, in each quadrant the effect of low error rate interacts with levels of pre achievement; effects tend to be greater at lower levels of pre achievement.

This interaction is reasonable if one considers low error rate to be an instructional treatment variable. As a treatment, low error rate would be positively associated with residual achievement because students learn more effectively when they spend more time working on tasks that they understand well (as indicated by their low error rate on those tasks). To a large extent, this is probably due to the fact that students in this study generally spent most of their time working independently (about two-thirds of all reading and mathematics instruction was conducted in the self-paced setting where there is very little direct instructional interaction). However, one would expect that students with greater skills and abilities in reading or mathematics would cope more effectively with new or more difficult material than would students at lower skill and ability levels. Therefore, these more capable students would learn reasonably well even if they received relatively little instruction at a low error rate.

This explanation fits the findings for all four quadrants in Table 4-16. Interestingly, the overall effect of low error rate appears to be smaller at grade 5, but the interaction is still pronounced, because high pre achievement students at grade 5 learn almost equally well at the different levels of low

error rate. It is possible that students increase their ability to deal with academic challenges as they acquire more experience in school.

This probing into the reasons for the positive association of low error rate and residual achievement is clearly speculative. However, for the purpose of generating hypotheses, the data appear to indicate that low error rate is an instructional treatment variable that is positively associated with student learning outcomes.

Summary: ALT and Student Learning

Generally consistent and strong effects were found in both the A-B and B-C data sets and across the four quadrants (reading and mathematics at grades 2 and 5). Allocated time, engagement rate, and low error rate were found to be positively related to student learning, while high error rate was negatively related.

It is interesting that ALT was most effective in predicting learning outcomes when the four components were analyzed as separate variables. The ALT model would propose that the simultaneous occurrence of allocated time, engagement, and low error should be more productive than the same duration of each occurring separately. For example, the ALT model suggests that allocated time without engagement should produce no learning. It should be remembered, however, that these four ALT variables were computed for this study in such a way that their simultaneous occurrence could not be determined in the form prescribed by the ALT model. That is, allocated time was measured for each of several specific content categories, but engagement rate, low error rate, and high error rate were computed over all reading or all mathematics instruction. These rates were computed over the more general content domains to insure satisfactory reliability (see Chapter 3). The result, however, is that engaged time with a low error rate in decoding blends

and long vowels was not computed directly. Indirect estimations of such simultaneous occurrences were found to be less powerful than the four separate variables analyzed jointly.

Despite the limitations of the present study, strong and surprisingly consistent effects have been shown for Academic Learning Time. Ongoing student learning behaviors have been successfully measured in the form of attentive involvement in a relevant instructional activity for which the student is able to respond correctly most of the time. The importance of these learning behaviors has been shown in reading and mathematics at grades 2 and 5. Measurement of these ongoing student learning behaviors has been shown to account for about one-tenth of the variance in student learning outcomes, as measured by achievement tests. Discussion has considered the likelihood that this proportion of variance would be increased considerably if measurement error of the achievement tests could be eliminated. Furthermore, the practical importance of the effects for ALT as they were measured has been illustrated with hypothetical examples. The need for caution in extrapolating these findings to extreme cases was also considered. Issues of causal inference were noted and special consideration was focused on inferences from the association of low error rate with learning.

CHAPTER 5: TEACHING PROCESS VARIABLES AND STUDENT ENGAGEMENT

From the analyses reported in the previous chapter, we know that student engagement rate is positively related to student performance on achievement tests. When students pay attention a greater percentage of time, they learn more. Knowing this, the next question is what do teachers do to increase student engagement rates. This chapter reports on two sets of analyses addressing this question. The first set of analyses examines the relationships between interactive teaching behaviors and student engagement. This section includes a discussion of instructional settings and the relationships between interactive teaching behavior, instructional setting, and engagement. The second set of analyses examines general characteristics of the instructional environment in relationship to student engagement.

INTERACTIVE TEACHING BEHAVIORS AND STUDENT ENGAGEMENT

The theoretical framework which guided the description of instruction in the BTES Phase III-B field study was one of the function or role of a behavior in the instructional process. Three interactive teaching functions are included in that model--presentation, monitoring and feedback. During direct interaction with the student in the classroom, the teacher presents information to the students, checks to see what the student knows or has learned, and provides feedback to the student about how he or she is doing. Behaviors representing these three functions were measured through extensive direct observation in the classroom. The BTES observation system was a time-based rotating-sample procedure. This classroom observation system recorded detailed information on the instructional day for individual target students. An event for each student was sampled about once every four minutes. Each event was coded along several dimensions. These dimensions included content of instruction, setting, student behavior (engagement and

error rate) and the instructional behavior provided by teachers and other instructors.¹

Classroom observation took place for a full day approximately once each week. Most data reported in this chapter are averages across about 15 days of observation in the B-C period (January to May).

The Interactive Teaching Behaviors

The interactive teaching behaviors that were observed are described in Table 5-1. Seven individual behaviors were coded. These behaviors represent aspects of the three interactive teaching functions. Notice that the behaviors can also be subdivided according to content. Some behaviors deal with substantive, academic content. Others deal with classroom procedure. For each of the seven behaviors, a description and an example are given.

All instruction was coded as occurring in one of two general instructional settings -- self-paced or other-paced. In self-paced settings the student works independently and determines for himself (within some time limit) the rate at which work gets done. The term "seatwork" is often used to describe activities of this type. In seatwork, each student works independently at his or her "seat" on some task. For instance, following reading group, the students may go back to their seats and do the workbook pages. Sometimes different students work on different tasks, as when students work on "contracts." While students are working in a self-paced setting, the teacher may be busy with other activities or may circulate

¹ The observation system also recorded the source of the instructional behavior (teacher, other adult, other student, or curriculum) and the focus of the instructional behavior (target student or group). Descriptive statistics on source and focus can be found in earlier technical notes (Filby, 1977; Filby and Fisher, 1977). No analyses of these two facets in relationship to other variables have been performed.

Table 5-1
Definitions and Examples of Interactive
Teaching Behaviors

General Function	Specific Behavior	Description	Example
SUBSTANTIVE BEHAVIORS			
Presentation	Explanation-Need	The instructor explains academic content in response to a clear and immediate student need for help.	During a seatwork assignment on short vowels, the teacher notices that Sonia is having trouble, so she goes over the sounds with her.
	Explanation-Planned	The instructor explains academic content but not in response to a specific or immediate student need. This often takes the form of a lecture.	The class comes to a section in the reading workbook on compound words. The teacher introduces this concept and puts some examples on the board.
Monitoring	Academic Observation	The instructor looks at or listens to student academic responses, in order to see how the student is doing.	While Eric is doing seatwork, the teacher looks at his paper to make sure he understands what to do.
	Academic Question	The instructor asks the student for a written or oral answer in order to observe and assess academic performance.	During group discussion, the teacher asks Mary to give the meaning of a word from that day's reading lesson.
Feedback	Academic Feedback	The instructor tells the student whether his answer is right or wrong. Could involve giving the right answer. Does not involve explanation.	The teacher asks Johnny to read out loud to the class the answers to the arithmetic problems. In oral reading circle, Susan hears Jim reading out loud while she reads along silently.
PROCEDURAL BEHAVIORS			
Presentation	Structure/Direct	The instructor states the goals of instruction or gives directions about the procedures and activities the student should carry out.	At the beginning of the math lesson, the teacher says "Today we are going to work on 10's and 1's like we did yesterday. Turn to page 20."
Feedback	Task Engagement Feedback	The instructor acts to control inappropriate behavior or to praise appropriate behavior. The focus is on the student's actions rather than the correctness of his answers.	Billy is out of his seat and standing by the pencil sharpener. The teacher tells him to "hurry up and get back to work."

around the room providing help as needed. One-to-one interaction between students or teacher and student may take place during a self-paced setting, but any group size larger than two changes the nature of the setting. Sometimes during a group lesson the teacher will let the students "finish the page on their own," allowing 1 to 5 minutes for independent work. The time students spent working individually would be coded as self-paced. In other-paced settings, the student must interact with other students or must share the teacher's attention with other students. The pace of instruction is determined by the whole interactive unit and is beyond the control of the individual student. The most frequent example of an other-paced setting is a group lesson, where the teacher instructs two or more students at the same time. A group of students working together without the teacher would also be coded other-paced.

Table 5-2 provides descriptive statistics on the percentage of the time that an individual student receives different interactive teaching behaviors in each of the two settings.² For instance, on the average, a student in grade 2 receives explanation based on need in a self-paced setting during 1.2% of his or her reading instruction. The sum of all behaviors with direct substantive academic content is shown in the rows labeled "Total Substantive." On the average, a student in grade 2 receives substantive instruction in a self-paced setting during 5.2% of his or her reading time.

To put these percentages in perspective overall statistics on percentage of time in each setting are also given. Students in both grades and both subject matter areas spend, on the average, at least 60% of their time in seatwork. There is more seatwork in grade 5 than in grade 2. There is more seatwork in mathematics than in reading. (This same pattern was reported in Phase II of BTES. See McDonald and Elias (1975)).

²More complete descriptive statistics, including class-level data, can be found in Filby and Fisher (1977).

Table 5-2

Frequency of Occurrence of Interactive Teaching Behaviors, by Setting

Teaching Behavior	Grade 2		Grade 5	
	Reading	Mathematics	Reading	Mathematics
Self-Paced Settings (Seatwork)				
Explanation-Need	1.2 (1.1)	2.2 (2.2)	0.4 (0.6)	1.5 (2.2)
Explanation-Planned	0.1 (0.3)	0.2 (0.6)	0.0 (0.2)	0.0 (0.2)
Academic Observation ⁺	0.2 (0.4)	0.3 (0.8)	0.1 (0.2)	0.0 (0.2)
Academic Question ⁺	0.3 (0.5)	0.1 (0.3)	0.1 (0.2)	0.1 (0.5)
Academic Feedback	3.4 (2.5)	2.4 (2.3)	1.2 (1.4)	1.1 (1.3)
Total Substantive [*]	5.2 (3.3)	5.2 (4.1)	1.9 (1.7)	2.7 (2.8)
Structure/Direct	0.4 (0.5)	0.4 (0.7)	0.1 (0.3)	0.1 (0.4)
Task Engagement Feedback	0.9 (1.2)	1.0 (1.6)	0.4 (0.8)	0.5 (0.8)
Total Interaction	6.5	6.6	2.3	3.3
No Teaching Behavior	57.8	66.9	66.4	72.7
Total	64.3	73.5	68.7	76.0
Other-Paced Settings (Group, Interactive)				
Explanation-Need	0.3 (0.4)	0.9 (1.5)	0.3 (0.4)	1.1 (1.6)
Explanation-Planned	1.9 (1.5)	3.3 (3.1)	2.1 (2.1)	4.5 (3.9)
Academic Observation ⁺	0.2 (0.4)	0.1 (0.3)	0.1 (0.2)	0.0 (0.0)
Academic Question ⁺	6.0 (3.2)	5.1 (4.8)	5.0 (3.2)	3.3 (3.9)
Academic Feedback	13.0 (6.3)	4.9 (4.1)	12.1 (5.8)	6.1 (5.8)
Total Substantive [*]	21.3 (8.2)	14.4 (9.7)	19.6 (7.6)	14.9 (10.6)
Structure/Direct	4.5 (2.6)	3.6 (3.2)	3.5 (2.2)	2.5 (2.3)
Task Engagement Feedback	0.8 (1.5)	0.6 (1.0)	0.5 (0.7)	0.6 (1.0)
Total Interaction	26.7	18.6	23.6	18.0
No Teaching Behavior	9.0	7.9	7.6	6.0
Total	35.7	26.5	31.2	24.0

Entries are: Mean (Standard Deviation)

All variables are expressed as percentages of student instructional time in reading or mathematics.

Sample size is 139 students in grade 2 and 122 students in grade 5.

Data came from the B-C period (January to May), with approximately 15 full days of observation per class.

⁺ Partly because academic observation was so infrequent, academic observation and academic question were combined for analysis. The combined variable was labelled "academic monitoring."^{*} Total Substantive includes Explanation-Need, Explanation-Planned, Academic Observation, Academic Question, and Academic Feedback.

The amount of instruction received by a student during seatwork is very small, especially in grade 5. Over half the time during reading and mathematics, students work on their own, with no instructional interaction.

Other-paced settings are, by definition, primarily interactive. Other-paced setting generally occur for less than one-third of the student's instructional time. During most, but not all, of this time some instructional interaction is occurring. Time with no immediate teaching behavior in a group setting would include things like having students put their names on their papers, having students briefly read some material or work a problem, or the teacher interacting with one student while the others wait.

There are some interesting patterns in the frequency of occurrence of different teaching behaviors in different settings and for different subject matter areas. Explanation occurs more often in mathematics than in reading. Apparently the content of mathematics lends itself more readily to the explanation of concepts and skills. Notice that explanation specifically based on immediate student need occurs more often in self-paced settings than in other-paced settings. One advantage to seatwork is that it allows the teacher to respond to individual students on a tutorial basis and to tailor the presentation to the student's needs.

In general, the most frequent behavior is academic feedback. This is true in both settings, grades, and subject matter areas. Academic feedback is more frequent in reading than in mathematics. In part, this is the reverse side of the trend toward more explanation in mathematics than in reading. It is also the case, however, that academic feedback includes a variety of different specific behaviors, some of which are unique to reading. In particular, oral reading involves high levels of feedback. When one student reads aloud while the target student reads silently, the

target student receives feedback from the student who is reading aloud; s/he hears the correct word as s/he attempts to read the word. (If the oral reader makes a mistake, then the teacher often intervenes with feedback.) Feedback during practice reading accounts for about half of the total academic feedback received in grade 2 reading and about one-third of the academic feedback received in grade 5 reading.

For purposes of this analysis, the percentage of the student's time for each behavior was summed across the two instructional settings so that we could focus attention first on differences between types of behavior. In addition, academic observation and academic question were combined to form the category "academic monitoring." This was done because the academic observation category was extremely infrequent. Means, standard deviations, reliabilities, and intercorrelations of these "main effect" variables collapsed across setting are shown in Tables 5-3 and 5-4.

The reliability coefficients that are reported estimate inter-observer reliability, that is, the extent to which two observers agree in coding a single day's events. The coefficients that are given estimate the reliability of an average of two observers, since two field workers were regularly assigned to each classroom.³ These estimates do not take into account variability across occasions. This is appropriate since engagement (and error rate) were measured on the same days as interactive teaching behaviors. (In effect, observation was continuous and no sampling of days is involved for analyses wholly within direct observation data.)

The reliability of measurement of these interactive teaching behaviors is not especially high. Interactive teaching behaviors occur infrequently,

³To calculate interobserver reliability, special paired-observation sessions were scheduled throughout the year (see Chapter 3). The actual data base for the study was collected by having the two regular field workers observe on alternate weeks.

Table 5-3

Means, Standard Deviations, Reliabilities and Intercorrelations of
Interactive Teaching Behaviors for the B-C Period
(Grade 2)

		1	2	3	4	5	6	Mathematics		Relia- bility
								Mean	SD	
Academic Monitoring	1	<i>.36</i>	.23	-.14	.54	.14	-.04	5.7	4.7	.74
Academic Feedback	2	.13	<i>.24</i>	-.09	.30	.05	.02	7.3	4.1	.72
Explanation-Need	3	-.05	.11	<i>.21</i>	.02	-.04	.13	3.2	2.7	.43
Explanation-Planned	4	.27	.39	-.01	<i>.32</i>	.02	.06	3.5	3.1	.44
Structure, Direct	5	.13	-.12	-.08	.09	<i>.59</i>	-.11	4.0	3.2	.36
Task Engagement Feedback	6	.03	.27	-.09	.10	.00	<i>.40</i>	1.6	1.8	.57
Reading - Mean		6.6	16.4	1.5	2.0	4.9	1.7			
SD		3.2	7.1	1.2	1.5	2.7	1.9			
Reliability		.61	.64	.48	.61	.48	.37			

Entries below the diagonal are for reading. Entries above the diagonal are for mathematics. Entries in the diagonal, in italic type, show the correlation between reading and mathematics.

All variables are expressed as percentages of student instructional time in reading or mathematics.

Sample size is 139 students.

Reliability coefficients are estimates of interobserver reliability based on 239 paired observations of individual students. The coefficients estimate the reliability of an average of two observers.

Table 5-4

Means, Standard Deviations, Reliabilities and Intercorrelations of
Interactive Teaching Behaviors for the B-C Period
(Grade 5)

		1	2	3	4	5	6	Mathematics		Relia-
								Mean	SD	bility
Academic Monitoring	1	<i>.33</i>	.35	-.03	.32	.26	.15	3.4	3.9	.60
Academic Feedback	2	.02	<i>.16</i>	.10	.22	.18	.02	7.1	5.7	.78
Explanation-Need	3	.21	.09	<i>.11</i>	-.04	-.13	.14	2.6	2.7	.68
Explanation-Planned	4	.30	.14	.12	<i>.33</i>	.30	.11	4.5	3.9	.69
Structure, Direct	5	.00	.28	.22	.11	<i>.27</i>	.18	2.7	2.3	.56
Task Engagement Feedback	6	.20	-.11	.06	.07	.11	<i>.37</i>	1.0	1.3	.48
Reading - Mean		5.3	13.2	0.8	2.2	3.6	0.9			
SD		3.3	5.9	0.8	2.1	2.2	1.1			
Reliability		.54	.89	.44	.51	.48	.84			

Entries below the diagonal are for reading. Entries above the diagonal are for mathematics. Entries in the diagonal, in italic type, show the correlation between reading and mathematics.

All variables are expressed as percentages of student instructional time in reading or mathematics.

Sample size is 122 students.

Reliability coefficients are estimates of interobserver reliability based on 201 paired observations of individual students. The coefficients estimate the reliability of an average of two observers.

from the perspective of a student's school day. Sampling fluctuations thus have a reasonably large effect on the data. It should be noted, however, that the variables used for relational analyses are averages over fifteen observation days. Some sampling errors should cancel out, thus increasing reliability above the estimates shown.

The intercorrelations indicate that there is some tendency for interactive behaviors to occur together. There are generally moderate positive correlations between academic monitoring, academic feedback, and explanation planned. This probably reflects the fact that some classes spend a fair amount of time in group lessons, where all these three behaviors tend to occur. Other classes have much less interaction of any kind.

Notice also that there is moderate consistency across subject matter areas. Students who receive more structuring and directing in reading also receive more in math. Even task engagement feedback shows consistency. In fact, it shows a high correlation considering the infrequency of the behavior and the moderate reliability. This is encouraging in that it suggests that sampling errors may cancel out over occasions, making the reported reliabilities a lower-bound estimate.

Correlations with Engagement

The relationships between interactive teaching behaviors and engagement were examined through regression and correlation. Analyses were run separately for the A-B and B-C periods. Commonality analysis by function and regression weights for each of the six behaviors in one multiple regression analysis can be found in Filby and Cahen (1977, 1978). Repeated here from the earlier reports are the correlations between each of the interactive behaviors

and engagement. The conclusions are similar regardless of analytic procedure. The correlations are given in Table 5-5.

Two sets of correlations are given in each of the four quadrants. The first set is correlations based on data from the B-C period (January to May), with approximately 15 full days of observations per class. The second set (in italic type) is correlations based on data from the A-B period (October to December) with about five days of observation per class. The data set from the B-C period is larger and variables are more reliably measured, therefore, B-C correlations should be regarded as the primary analysis. Correlations from the A-B period provide a form of replication.

For these analyses, engagement rate was defined within time period. Engagement rate was defined as the percentage of observation time when the student appeared to be attending to the instructional task. The reliability of the engagement rate variable during the B-C period was about .70 in grade 2 and .77 in grade 5.

The results for individual variables, shown in Table 5-5, will be discussed by function, beginning with monitoring.

Academic monitoring is represented by one variable, a combination of the observation categories of academic observation and academic question. Primarily, this variable indicates the frequency with which the teacher asks questions of students. Frequency of monitoring is positively correlated with engagement in all four quadrants. The more often a student is asked a question, the higher his engagement rate. This result was particularly consistent across both quadrants and time periods.

Feedback is represented by two variables with very different patterns of association. Academic feedback is information to the student about the correctness of his academic responses. In the A-B period, academic feedback

Table 5-5

Correlations of Interactive
Teaching Behaviors with % Engaged

General Function	Specific Behavior	Quadrant ³			
		2R	2M	5R	5M
Presentation	Explanation-Need	04 08 ¹	-02 -03	09 -13	-15 -10
	Explanation-Planned	24 14	27 01	09 02	45 25
	Structure, Direct	-03 -13	20 15	-01 -04	29 32
Monitoring	Academic Monitoring	37 22	40 23	19 26	31 19
Feedback	Academic Feedback	39 45	29 11	42 14	39 28
	Task Engagement Feedback	-22 -14	-25 -22	-43 -23	-30 -06
Total Substantive Interaction ²		48	42	43	46

¹The numbers in gothic type are the correlations obtained in the B-C period (January to May), with approximately 15 full days of observation per class. The numbers in *italic* type are corresponding correlations from the A-B period (October to December), with approximately 5 days of observation per class.

²Total Substantive is the sum of four behaviors -- Explanation-Need, Explanation-Planned, Academic Monitoring, and Academic Feedback.

³The general term "quadrant" is used to describe any one of the four sets of data: grade 2 reading (2R), grade 2 mathematics (2M), grade 5 reading (5R), and grade 5 mathematics (5M). The same students were followed in reading and mathematics at each grade.

Sample size is 139 students in grade 2 and 122 students in grade 5. Correlations reach significance at the .05 level when $r \geq .18$.

had an especially high positive correlation with engagement in grade 2 reading. In the B-C data set, the effect is more general. In all four quadrants, academic feedback has a significant, positive correlation with engagement. Students who receive more academic feedback have higher rates of engagement. This effect is particularly strong in reading.

Task engagement feedback is consistently negatively correlated with engagement. Students who receive more task engagement feedback have lower rates of engagement. Task engagement feedback gives the student information about the appropriateness of his or her behavior. Theoretically, it includes both positive feedback ("I like the way you're working") and negative feedback ("Pay attention or you'll miss recess"). An examination of the A-B results suggested the inference that task engagement feedback is primarily negative and that it is elicited by lack of student engagement. When students are working smoothly, engagement is high and relatively little engagement feedback is given (not much praise). When students are not working smoothly and off-task behavior occurs, the teacher responds to control the inappropriate behavior. Some support for this hypothesis was found in a breakdown of the B-C data. Two combination variables were created: task engagement feedback in connection with student on-task behavior ("positive feedback") and task engagement feedback in connection with off-task behavior ("negative feedback"). Positive feedback was extremely rare. On the average, in each quadrant, it occurred less than half of one percent of the student's instructional time. It accounted for only about 20% of the total task engagement feedback in grade 2 and only 10% of the total in grade 5. Task engagement feedback is primarily negative.

These results for task engagement feedback are a striking counterpoint to the general pattern of positive correlations for interactive teaching

behaviors and are worth some discussion. The complexity of interpreting correlational data and inferring causal relationships is particularly apparent here. Intuitively it might seem that task engagement feedback, even negative feedback, is intended to increase student engagement. The teacher reminds a student to get to work in order to stop off-task behavior and start on-task behavior. Seemingly, exactly the opposite occurs. Students who receive more task engagement feedback are students who are off-task more often. One interpretation, consistent with the data but based on inference, is that some students are more inclined to goof-off than others and that these students attract the teacher's attention when things get too bad and thus elicit task engagement feedback.

The presentation variables appear to be more strongly related to engagement in mathematics than in reading. Structure/direct involves presentation of procedure, either a description of the content to be covered or specific directions on what to do. Structure/direct has a generally positive relationship with engagement in mathematics. This category accounts uniquely for only 1% or 2% of the variance in engagement, but the pattern is consistent across both grades and time periods. It appears that, in mathematics, students who receive more structuring and direction from the teacher have higher rates of engagement.

Explanation-planned also has significant positive correlations with engagement in mathematics. The reader may recall from Table 5.2 that explanation, unlike other variables, occurs relatively more frequently in mathematics than in reading. Planned explanation of concepts and procedures seems to be called for more often in mathematics than in reading, and its occurrence is positively associated with engagement.

These results for three interactive teaching functions and the individual behaviors can be summarized in a single parsimonious explanation of the data -- the more often a student is involved in academic interaction with an instructor, the higher his engagement. As a type of statistical summary, we computed the correlations between total substantive interaction and percent engaged. This correlation, for each quadrant, is shown at the bottom of Table 5.5. Total substantive interaction is the sum of four behaviors -- explanation-need, explanation-planned, academic monitoring and academic feedback. This particular combination was used because it was developed in the theoretical grouping of variables. Note both the size of the correlation and the consistency across quadrants. The strongest effect is at the general level -- more interaction is associated with more engagement.

Teaching Behavior, Setting, and Engagement

From the descriptive statistics in Table 5-2 we know that most instructional interaction takes place in a group setting. From the correlations in Table 5-5 we know that the amount of instructional interaction is positively associated with student engagement rate. One way of integrating these two facts is to state that setting is related to engagement rate. Engagement rates are higher in a group setting than during independent seatwork. The more time a student spends in seatwork, the lower his overall engagement rate is likely to be. One estimation of the magnitude to this effect is shown in Table 5-6.

The average engagement rate in self-paced settings like seatwork is at least 13% lower than the average engagement rate in other-paced groupwork. If some

Table 5-6

**Variability in Use of Instructional Settings
and Average Engagement Rate, by Setting**

Grade	Subject	Setting	<u>% of Time Spent in Setting</u>			Average Engagement Rate
			Average	Minimum	Maximum	
2	Reading	Self-Paced	64%	43%	91%	68%
		Other-Paced	36%	9%	57%	84%
2	Mathematics	Self-Paced	73%	22%	96%	67%
		Other Paced	27%	4%	78%	82%
5	Reading	Self-Paced	69%	37%	92%	70%
		Other Paced	31%	8%	63%	84%
5	Mathematics	Self-Paced	76%	37%	100%	72%
		Other-Paced	24%	0%	63%	85%

Sample size is 139 students in grade 2 and 122 students in grade 5.

Data came from the B-C period (January to May) with approximately 15 full days of observation per class.

students spend a much higher proportion of time in seatwork than others, a lower engagement rate is likely to result. Looking at the columns on minimum and maximum proportion of time in each setting, it appears that there are dramatic differences in use of different settings. In grade two reading, one student may spend 91% of his or her time in seatwork while another student spends only 43% of his or her time in seatwork. This difference in setting allocation translates into a 7% difference in overall engagement rate, based on average engagement rates in Table 5-6.⁴

This setting effect makes practical sense. Students are more likely to goof-off when they are on their own, with only general supervision by the teacher, than they are when they are in a group with direct teacher attention.

The setting effect is only part, however, of a more general positive effect of academic interaction between instructor and student. The average correlation between total substantive interaction and engagement, as shown in Table 5-5, is about .45. The average correlation between percentage of time in groupwork and engagement is only about .31. "Total substantive interaction" is a better description of the desired condition than is "groupwork."

There are three reasons why instructional interaction shows a positive effect over and above setting. Two have to do with the fact that time in a group setting includes time when no instructional interaction is taking

⁴ Actually, this difference is underestimated, because the average engagement rates on Table 5-6 do not take into account an additional effect of setting on engagement. Engagement is, on the average, lower in seatwork than in groupwork. Moreover, the higher the percentage of time in seatwork, the lower the engagement rate in seatwork! That is, the class with 43% seatwork is really likely to have an engagement rate of 72% in seatwork. The class with 91% seatwork is really likely to have an engagement rate of 60% in seatwork. This makes the overall difference in engagement rate more than 16% for these two extremes.

place and during this time engagement is very low. As show in Table 5.3 about one-quarter of the time during groupwork there is no immediate instructional behavior. But this percentage varies from one class to another. While substantive instructional interaction is taking place during groupwork, the level of engagement is very high. The range is from 95% engaged during substantive instructional behavior in grade five reading to 98% engaged during substantive instructional behavior in the other three quadrants. While no instructional behavior is taking place in groupwork, the engagement rate drops to around 50%. (The range is from 48% to 53% across quadrants.) What happens during this time? For one thing, the teacher waits while the students put their names on their papers. Even while the student was writing his or her name s/he would be coded "not engaged." This would be labeled an "interim" task, and the student would be coded "not engaged-interim." Although s/he may be doing what s/he is supposed to do, the student is not learning math or reading while s/he writes his or her name. Another type of not-engaged activity during groupwork is waiting. Some students wait while other students finish writing their names or finish working a problem. This reflects one of the disadvantages of other-paced settings -- the student cannot go ahead at his or her own rate. In summary, engagement is high in group settings while instructional behavior is occurring and low in group settings when no immediate instructional behavior is occurring.

Naturally, the overall engagement rate in group settings is highly influenced by the amount of time when no instructional behavior is occurring. The percentage of other-paced setting time with no instructional behavior ranges from 0% to over 70%. The impact of this difference is reflected in the correlations in Table 5.7.

Table 5-7

**Correlations Between Instructional Behavior and Engagement
Within and Across Settings**

Interaction Variable	Engagement Variable	Quadrant			
		2R	2M	5R	5M
Percent of Other-Paced Time During Which Substantive Interaction Takes Place	Engagement in Other-Paced Settings	.54	.46	.56	.47
Percent of Other-Paced Time During Which Substantive Interaction Takes Place	Engagement in Self-Paced Settings	.31	.03	.46	.11
Percent of Self-Paced Time During Which Substantive Interaction Takes Place	Engagement in Self-Paced Settings	.18	.15	.18	-.14

Sample size is 139 students in grade 2 and 122 students in grade 5. A correlation $\geq .18$ is significant at the .05 level.

Correlations are based on data from the B-C period.

Substantive interaction is the sum of explanation-planned, explanation-need, academic monitoring, and academic feedback.

In all four quadrants there is a high correlation between the percent of other-paced time when instructional behavior is occurring and the engagement rate in other-paced settings.

Notice also in row two of Table 5-7 that there is a degree of transfer in the impact of instructional behavior in other-paced settings. The percentage of other-paced time with instructional behavior is related to engagement rates in seatwork. One interpretation is that a pacing or momentum effect is involved. If during group lessons most of the time is directly involved with substantive instruction this may contribute to a sense of purposefulness. Students will expect to work and will apply the same efficiency to seatwork. They will sit down and go to work. A smaller proportion of actual academic instruction may be associated with a sense of slow pacing and lack of real emphasis on academic learning. This general atmosphere may encourage off-task behavior during seatwork.

This effect of pacing on engagement in both other-paced and self-paced settings appears to be related to a finding reported by Ogden, Brophy, and Evertson (1977) in their study of organization and management techniques in first grade reading groups. They found that reading achievement was strongly related to the time lapse between teacher and children reaching the reading group and the actual beginning of the lesson. They concluded that "Once the last person (teacher or child) arrives in a reading group, the teacher should begin the lesson immediately. Any materials necessary for the lesson should be pre-arranged so that reading group time is not wasted on organizational tasks" (Ogden, Brophy, and Evertson, 1977,

page 14). Pacing appears to relate to achievement, with student engagement rate as a mediating factor.

There is one final way in which teacher-student interaction influences engagement beyond the role of group settings. During seatwork, engagement rates are generally lower than in group settings, but engagement rates are higher during seatwork when students get more contact with the teacher. During seatwork some teachers move around the room giving feedback and explanation; other teachers are busy doing other things. With a number of children to be covered, any one child cannot receive a lot of teacher attention; but in the classes we observed some students received substantive instruction during as much as one-fourth of the time they were doing seatwork while other students never received individual attention during seatwork.

The correlations in Table 5-7 indicate that in three of the four quadrants the amount of substantive instruction a child received during seatwork was positively associated with student engagement in seatwork. This effect is even greater when one looks at only those classes with a high percentage of time in seatwork. As a group, these classes tend to have lower engagement rates and somewhat higher percentages of substantive instruction during seatwork. Within this group, the amount of substantive instruction in seatwork tends to be positively related to engagement. In grade two math, classes with over 70% seatwork were divided into two groups. One group had substantive interaction during a relatively large percent of seatwork time (an average of 11%); the other group had sub-

substantive instruction during only about 5% of seatwork time. The engagement rate during seatwork in the high interaction classes averaged 71%. The engagement rate during seatwork in the low interaction classes averaged 61%. In other words, if for some reason a large percentage of instructional time is spent in seatwork, it is important to monitor the students and provide academic feedback and explanation as often as possible.

This section on setting can be summarized in three inferred recommendations for instruction:

1. Part of the instructional time should be spent in a group lesson. (Engagement rates are higher in group settings. As the percentage of time in seatwork increases, the engagement rate during seatwork).
2. During group lessons, as much of the time as possible should be spent actually working on substantive content. The amount of time spent in organizational activities or waiting for other students should be minimized. (This brisk pacing with a focus on academic content is associated with higher engagement rates in group settings and during seatwork.)
3. During seatwork, especially when a large percentage of time is spent in seatwork, the teacher should monitor student activities and provide academic feedback and/or explanation. (Engagement rates are higher in seatwork when students have more contact with an instructor.)

Within-Student Analysis

In the preceding analyses, information about each student was averaged across all observation days in the period. Analyses based on student averages allow us to say that a student who receives more substantive interaction than some other student tends, on the average, to have a higher engagement rate than that other student. But we don't really know anything about how behaviors impact on a single student over time. It is possible that effects would be different within student from between students.

One place where differences might occur is for task engagement feedback. We know that a student who is off-task more often is more likely to receive task engagement feedback than a student who is more attentive. This suggests that task engagement feedback is not an effective form of management. It is possible, however, that some students are prone to be off-task; but on days when the teacher watches them more closely and prompts them to get to work, they pay attention more. This would lead to a positive correlation within student across days between task engagement feedback and student engagement.

In order to examine this possibility we calculated a pooled, within-student correlation. We started with a set of data giving engagement rate and task engagement feedback rate for each student on each observation day in the B-C period (over 1700 data points in grade two). We then mean-deviated each observation from the student's average score for the variable. For instance, if a student had an engagement rate of 80% in reading on a particular day and his or her average engagement rate over the B-C period was only 70%, then his or her mean-deviated score for that day would be +10%. That student had higher-than-average engagement on that day. Having carried out this procedure on both variables, we correlated the two mean-deviated scores.

The results were very similar across quadrants and will be summarized here. Between students, the correlation between task engagement feedback and engagement averaged $-.30$, as reported in Table 5-5. Within student, the correlation averaged $-.20$. The relationship is in the same direction and still significant, although reduced in magnitude. That is, a student gets more task engagement feedback on a day when s/he is off-task more often. The difference in magnitude between the two correlations suggests,

however, that students who are generally off-task sometimes get task-engagement feedback on days when they are relatively highly engaged. This may mean some positive benefit of task engagement feedback. It may also mean that teachers single out some misbehaving children as targets for complaint regardless of whether it is particularly appropriate for that child on that day.

One illustration of the experience of a student over time is given in Figure 5-1. This figure plots engagement rate and task engagement feedback across observation occasions in the B-C period for a boy in grade two. The date of each observation is given on the horizontal axis. There are no entries for 1-20 and 2-16 because the class had no math on 1-20 and the student was absent on 2-16. This figure shows trends in engagement that seem to be related to trends in task engagement feedback. In particular, the first three data points show a period during which the student showed a declining engagement rate and drew increasing amounts of task engagement feedback. After 2-01, student engagement went up and task engagement feedback dropped off. Notice that the student again drew task engagement feedback on 3-22, and engagement was relatively high for this student on that day. These specific instances seem to match the general trends evident in the correlational data. There is not much support for the belief that task engagement feedback has a facilitory effect.

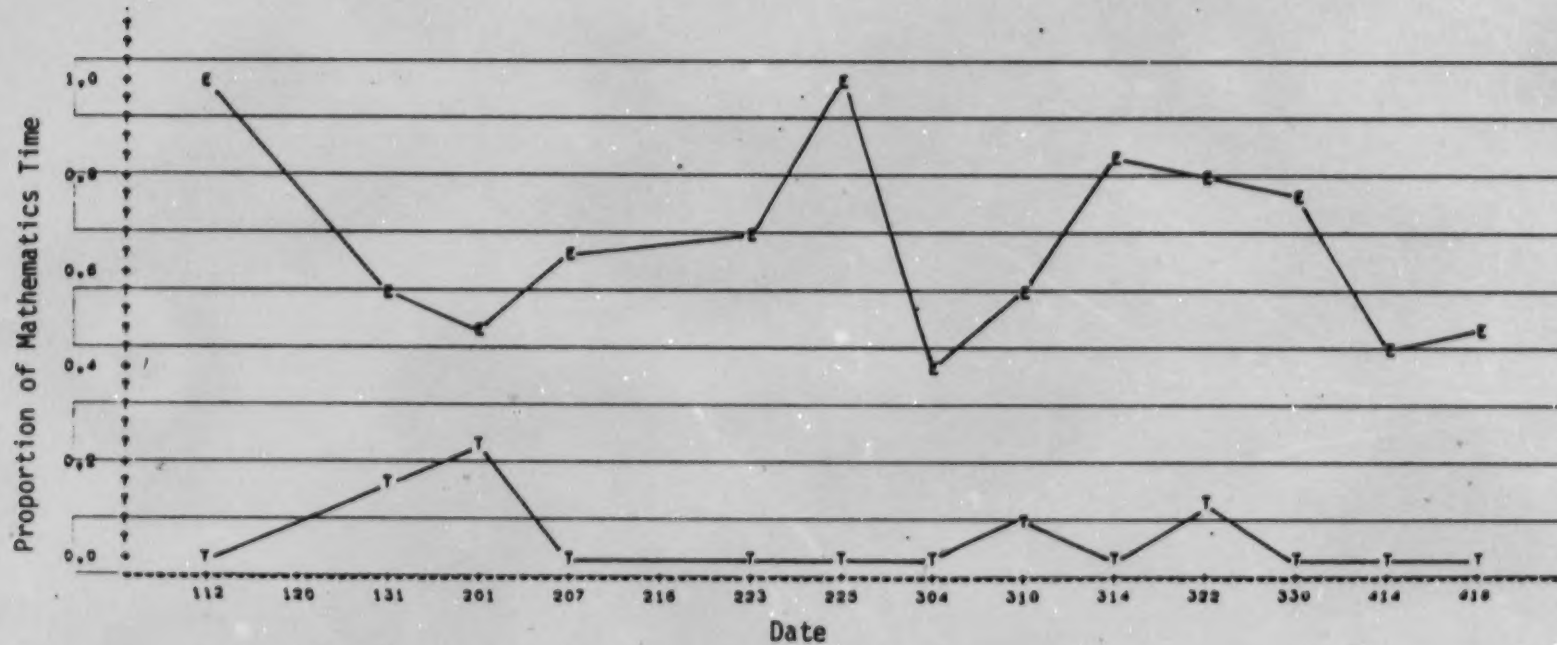
GENERAL CLASSROOM VARIABLES AND STUDENT ENGAGEMENT

The General Variables

In addition to measures representing each of the five teaching functions (diagnosis, prescription, presentation, monitoring, and feedback), other information about BTES classrooms was collected. These measures

Figure 5-1

Plot of Engagement Rate and Task Engagement Feedback
Across Weeks for a Single Student



E = Engagement Rate

T = Task Engagement Feedback Rate

Student is a boy in grade two.

5-25

describe general characteristics of the classroom, such as general characteristics of the instructional program or aspects of the classroom atmosphere or learning environment.

These general measures come from ratings made by the field workers who carried out observation in each classroom. Two types of ratings were obtained. Following each day's observation, the field workers filled out 15 rating scales describing aspects of that particular day. Since observation was carried out about 5 times during the A-B period and 15 times in the B-C period for each class, we had about that same number of daily ratings on each class. These ratings were averaged within time period to get an overall estimate of the class's standing on each of the 15 scales.

In addition to these daily ratings, field workers completed another set of ratings in the spring. Some of these rating scales were designed to get at aspects of diagnosis and prescription. For this set of ratings field workers reviewed all protocols of interviews with their regularly assigned teachers. Based on this interview data as well as their classroom observation experience, the field workers completed 11 rating scales. Since two field workers were regularly assigned to each classroom, we had two independent ratings on each class. This allowed us to estimate the reliability of the ratings. For relational analysis, we averaged the two scores on each scale.

These 26 ratings along with measures from other sources, such as tests, questionnaires, and interviews, comprised a large set of measures of diagnosis, prescription, and general characteristics. In fact, the total number of measures obtained was larger than the number of classes in the study, a mind-boggling task for quantitative analysis. This was done because we anticipated that not all measures would prove reliable and

because we wanted multiple opportunities to measure important variables. Nevertheless, it was a major undertaking to reduce the size of the variable pool in a manner that was satisfying to both the logic of the measures and the empirical evidence of interrelationships. The report on A-B analyses (Technical Note V-1b, Filby and Cahen, 1977) includes a detailed description of the process of variable reduction and definition along with a complete set of descriptive statistics.

In this process of variable reduction and definition, five general variables were created. These variables are described in Table 5.8.

The first of the five variables evaluates the appropriateness of the instructional program or curriculum relative to the needs of individual students. One could have considered this variable to be within the domain of the prescription function. It does evaluate the quality of the prescription decisions that are made. One would hope that prescription would build on diagnosis and would represent a reasonable treatment suited to the individual student. In the absence of good empirical evidence on the efficacy of different treatments for different students (the content of prescription), it is impossible to judge quality objectively. We attempted, however, to get a rating from the field workers on one aspect of the perceived quality of prescription "the appropriateness of instruction to student needs." This variable emphasizes the individualization and flexibility of the program. It asks the field workers to rate "the extent to which the instructional program is responsive to the individual needs and level of readiness of the student." This rating of appropriateness was not included among the prescription variables because it reflects a combination of diagnosis and prescription. The variable is intended to estimate the extent to which knowledge of students is acted upon to

Table 5-8

Definitions of the General Variables

Variable Label	Method of Data Collection	Description
A, AUN Appropriateness of Instruction to Student Needs	Single rating based on interviews and observation (7-point scale)	Field workers were asked to rate "the extent to which the instructional program is appropriate to the level of readiness and needs of students." This includes the appropriateness of task difficulty and flexibility of pacing to allow for differential rates of learning.
G1 Global Rating (Like Class?)	Single rating based on interviews and observation (7-point scale)	Field workers responded to the question "Would you like your (hypothetical) child to be in this teacher's class?"
G2, G2UN Perceived Academic Competence	A combination of 9 rating scales completed after each observation (each a 7-point scale; unit-weight combination)	Nine intercorrelated ratings were averaged to get an estimate of the field workers' perception of the quality of the academic instruction: Cognitive Task Orientation Clarity Structuring Knowledge of the Pedagogy of Reading and Math Complexity of Communication Consistency of Message Attention to Students Expectation of Student Success
G3, G3UN Orientation Toward Affect	A combination of 3 rating scales completed after each observation and 2 ratings based on interviews and observation (each a 7-point scale; unit-weight combination within method and then across)	These intercorrelated ratings describe the classroom climate and the degree to which the teacher is aware of, encourages, and values the affective domain. Daily ratings: Conviviality Acceptance of Feelings Flexibility Interview-based ratings: Awareness of Affect Value Placed on Affective Outcomes
G4, G4UN Learning Environment	A combination of 2 rating scales completed after each observation (each a 7-point scale; unit-weight combination)	These intercorrelated ratings describe the extent to which students take responsibility for classroom activities and work cooperatively with each other: Cooperation Student Responsibility

prescribe apparently reasonable instructional programs. Because of this factor, appropriateness was included as a general variable.

A central variable in the set of general variables is the global rating, G1. This is a rating the field workers made in response to the question, "Would you like your (hypothetical) child to be in this class?" We asked this question in order to get the field worker's summary judgment of each class. On the one hand, this summary judgment is a meaningful predictor variable, especially in initial screening analysis to locate relationships to examine in more detail. On the other hand, this summary judgment provides a measure of "halo," global liking or disliking, which can be used (and was used) to "correct" other ratings. This is a complex issue which needs to be discussed in some detail.

Residual variables. The global rating summarizes the field worker's judgment of the class. In one sense it is "halo," a global impression of liking or disliking. It is so general that it cannot lead to specific knowledge about the components of quality instruction. But it also represents a complex and reliable evaluation of the interrelationships of many factors. Educational research represents an attempt to isolate and define the components of good education using logic and analysis. The human mind is incredibly skilled at integrating many factors to achieve intuitive understanding. The global rating represents this type of complex judgment. Research struggles to explain the sources of this type of judgment.

Both appropriateness and the other general variables are highly correlated with the global rating, G1. This happens for two reasons. One is that these different variables actually represent aspects of a favorable educational environment. People want their children to be in classes that

are truly judged to be appropriate, academically competent, warm, and cooperative/responsible. The second reason is highly related to the first and has to do with the fact that these ratings were not carried out independently. Because people value these characteristics, they tend to give higher ratings to classes they like. (If I like teacher X and her classroom and if academic competence is part of good education, teacher X must be competent.) In other words, each of the general variables contains a part which is realistic judgment of a particular dimension and a part which is "halo."

In order to examine the effect of general variables independent of G1, we created a residual version of each of the variables, partialling out G1. This residual variable was derived by regressing a general variable on G1 and getting the difference between the observed score and the score you would predict on the basis of G1 alone. The original, unresidualized variable always has the letters "UN" in the label. For instance, AUN is the rating on appropriateness of instruction which was discussed above. The residual score for appropriateness, labeled "A," represents deviation score from the regression line predicting AUN with G1. The residual score for a particular rating isolates that part of the rating which is separate from the global rating. It identifies a class which is particularly high (or low) on a particular dimension relative to that class's level on the general mix of variables. Where "halo" is a real problem in the ratings, the residual score is a better measure of the labeled dimension. As will be seen in the results section, prediction of outcomes can be understood best by looking at the relationship of engagement with both residualized and unresidualized scores.

Besides appropriateness and the global rating, there are three other

general variables -- perceived academic competence, orientation toward affect, and learning environment. These variables are described in Table 5-8.

Descriptive Statistics

Means, standard deviations, and reliabilities of the general variables are given in Table 5-9. Each variable was transformed to a mean of 50 and a standard deviation of 10. This transformation does not change the shape of the underlying distribution or the magnitude of a variable's correlation with an outcome. Variables were standardized because there is no absolute meaning of a specific score on a specific scale. Standardizing made it easier to compare a teacher's relative position on one measure with his or her relative position on another measure.

The reliability of general variables was in most cases fairly high. Residual scores contain a higher proportion of error than unresidualized scores. Details on calculation of reliability can be found in Technical Note V-1b (Filby and Cahen, 1977).

Intercorrelations of general variables are given in Tables 5-10 and 5-11. These correlations are for the B-C period. All of the unresidualized variables show high positive intercorrelations. The residualized variables, as expected, show somewhat lower intercorrelations.

It is particularly interesting to look at the correlations of G2 with G3 and G2UN with G3UN. In grade 2, G2UN and G3UN correlate .66. The residualized variables G2 and G3 correlate .41. In grade 5, these two dimensions are more distinct. G2UN and G3UN correlate .41. The residualized variables G2 and G3 correlate -.19. While both dimensions, academic competence and orientation toward affect, contribute to a favorable overall rating in the eyes of the field workers, they contribute relatively

Table 5-9
Means, Standard Deviations and
Reliabilities of the General Variables
in the B-C Period

	Variable	Mean	Standard Deviation	Reliability			
				Grade 2 Rdg. Math	Grade 5 Rdg. Math	Grade 2 Rdg. Math	Grade 5 Rdg. Math
A	Appropriateness of Instruction (Responsiveness to Individual Needs)-Residualized	50	10	.60	.60	.50	.71
AUN	Appropriateness of Instruction	50	10	.72	.75	.74	.81
G1	Global Rating ("Like Class?")	50	10	.80		.85	
G2	Perceived Academic Competence-Residualized	50	10	.84		.89	
G2UN	Perceived Academic Competence	50	10	.91		.95	
G3	Orientation Toward Affect-Residualized	50	10	.65		.57	
G3UN	Orientation Toward Affect	50	10	.75		.70	
G4	Learning Environment (Cooperation, Student Responsibility)-Residualized	50	10	*		*	
G4UN	Learning Environment	50	10	.91		.87	

* Unavailable

Reliability of A, G1 and G4 variables is inter-rater reliability, based on two ratings of each class. Reliability of G2 and G3 variables is the reliability of a composite.

Note: Variables A, AUN, and G1 are the same as those used in A-B analyses. The remaining variables are composites containing elements which are ratings on B-C observation occasions; they are therefore not identical with corresponding A-B variables.

Table 5-10

Intercorrelations of General Variables
(Grade 2, B-C Period)

	AR	ARUN	AM	AMUN	G2	G2UN	G3	G3UN	G4	G4UN
ARUN	73									
AM	57	41								
AMUN	43	76	76							
G2	59	43	31	24						
G2UN	45	77	24	61	76					
G3	33	24	12	09	41	30				
G3UN	25	63	09	50	31	66	75			
G4	12	09	-12	-09	34	25	02	02		
G4UN	10	50	-10	34	27	60	01	43	78	
G1	00	68	00	65	00	65	00	66	00	63

Sample size is 25 classes.

Table 5-11
Intercorrelations of General Variables
(Grade 5, B-C Period)

	AR	ARUN	AM	AMUN	G2	G2UN	G3	G3UN	G4	G4UN
ARUN	69									
AM	71	49								
AMUN	57	82	81							
G2	58	40	49	39						
G2UN	39	81	33	70	67					
G3	08	05	-06	-05	-19	-13				
G3UN	06	53	-04	37	-14	41	74			
G4	56	38	44	35	72	48	-04	-02		
G4UN	46	73	36	63	58	82	-03	37	82	
G1	00	73	00	59	00	74	00	67	00	58

Sample size is 21 classes.

independently in grade five.

Two more trends are apparent in Tables 5-10 and 5-11. One is that appropriateness of instruction, adaptation to the needs of individual students, is highly correlated with perceived academic competence. This is particularly true for appropriateness in reading. The second trend is that a cooperative learning environment is very highly correlated with perceived academic competence in grade 5. Even the residualized variables correlate .72 (compared with .36 in the A-B period). The correlation is lower in grade 2 and slightly lower in the A-B period than in the B-C period. Apparently by the end of the year in grade 5 cooperation and student responsibility are a strong component of the field workers' image of an academically competent classroom.

Correlations with Engagement

We have seen that interactive teaching behaviors are related to student engagement rate. It is also possible that general characteristics of the classroom atmosphere or environment are associated with engagement rate. To examine this possibility, we calculated correlations between each of the general variables and engagement rate. Because the general variables, unlike interactive teaching behaviors, are all measured at the classroom level, class mean engagement was calculated and all analyses were conducted at the class level. The variable of class mean engagement was found to be consistent across the A-B and B-C periods ($r = .88$ in grade 5).

Because of the high degree of intercorrelation among the variables, no regression analyses were performed. Instead, we used the global rating, G1 for an initial screening analysis by looking at the correlation between G1 and engagement. Table 5 12 reports this correlation, as well as the correlations of the other general variables and engagement, for each of the four quadrants in the study. Correlations from the B-C period are

Table 5-12
Correlations of General Variables
with % Engaged

Variable	Grade 2		Grade 5	
	Rdg.	Math	Rdg.	Math
G1 Global Rating ("Like Class?")	07 <i>-03</i>	02 <i>00</i>	62 <i>56</i>	79 <i>61</i>
G2 Academic Competence-Residualized	07 <i>11</i>	-09 <i>12</i>	22 <i>-05</i>	18 <i>07</i>
G2UN Academic Competence	11 <i>06</i>	-05 <i>10</i>	60 <i>39</i>	71 <i>51</i>
G3 Orientation Toward Affect-Residualized	-28 <i>-09</i>	-33 <i>-27</i>	-55 <i>-61</i>	-38 <i>-35</i>
G3UN Orientation Toward Affect	-17 <i>-17</i>	-23 <i>-27</i>	01 <i>-24</i>	26 <i>-01</i>
G4 Learning Environment-Residualized	-10 <i>-10</i>	02 <i>-06</i>	32 <i>22</i>	14 <i>11</i>
G4UN Learning Environment	-04 <i>-10</i>	03 <i>-05</i>	62 <i>40</i>	58 <i>32</i>
A Appropriateness of Instruction-Residualized	13 <i>30</i>	07 <i>00</i>	20 <i>-09</i>	-14 <i>-21</i>
AUN Appropriateness of Instruction	14 <i>20</i>	07 <i>00</i>	58 <i>35</i>	35 <i>20</i>

The numbers in italics are the correlations obtained in the A-B period.

Sample size is 25 classes in grade 2 and 21 classes in grade 5. Correlations reach significance at or beyond the .05 level when $r > .45$. (The level of significance will differ somewhat as a function of sample size and slightly due to residualized versus non-residualized definitions.)

given in dark type. Corresponding A-B values are shown in italic type.

The pattern of results is very similar for both time periods. In both sets of data there is a strong difference between grade levels. In grade 2 the global rating is unrelated to engagement. In grade 5 there is a strong positive relationship; classes with more positive global ratings by field observers have higher rates of engagement than classes with less positive ratings.

As a follow-up, we looked at the correlations of the other general variables with engagement. Following the logic of G1 as a screening analysis, one might have stopped with G1 in grade 2. One would expect that variables highly correlated with G1, that is, the four unresidualized general variables, would not correlate significantly with engagement if G1 itself does not correlate significantly. To test this expectation, and out of curiosity about how residual variables would correlate, we looked at all correlations for grade 2. None of the correlations were statistically significant at the $p \leq .05$ level. Residualized and unresidualized versions of the same variables had about the same correlations with engagement in grade 2. The strongest effect was a trend toward a negative relationship between engagement and orientation toward affect.

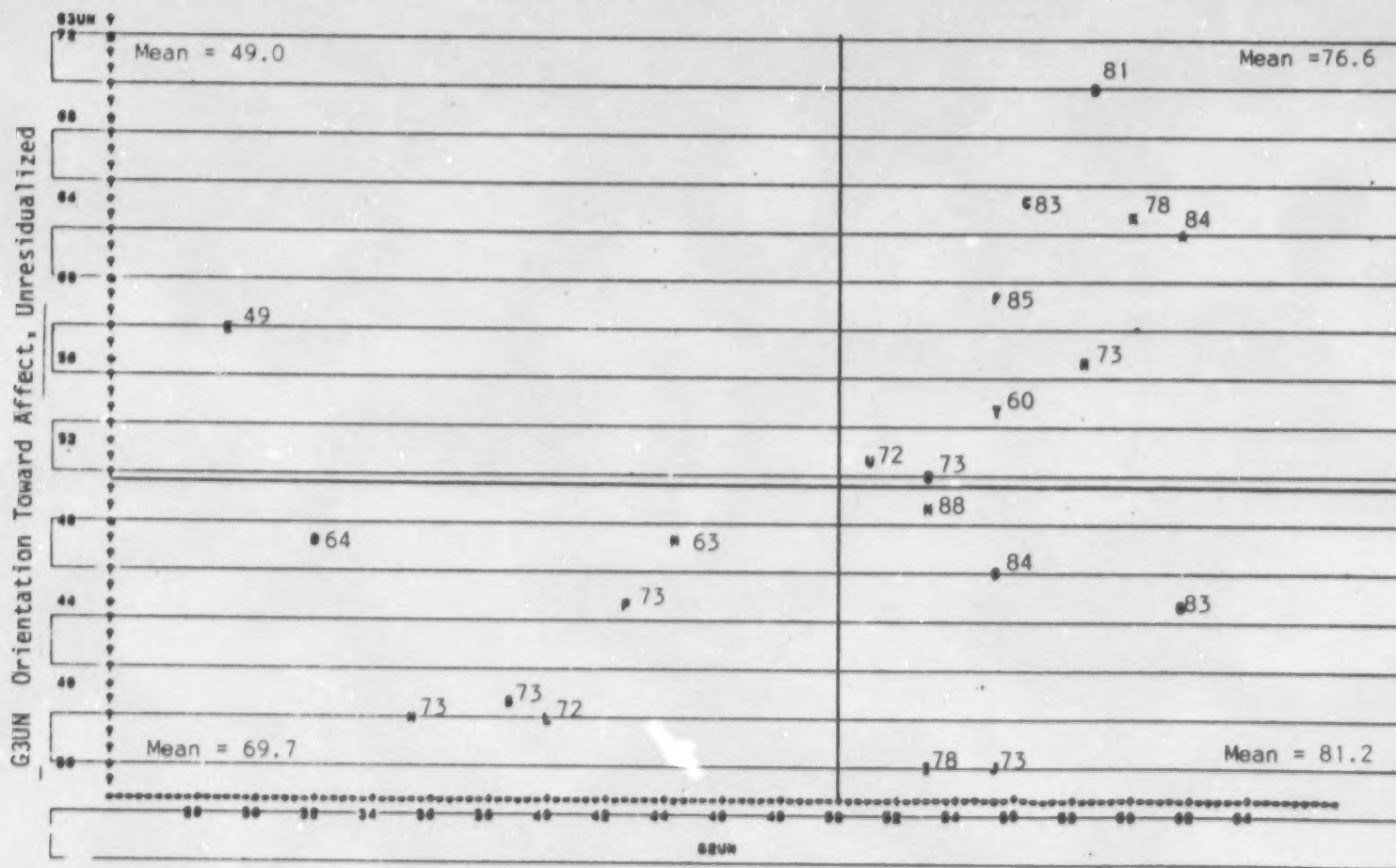
In grade 5, a complex pattern of interrelationships emerged. For the unresidualized variables, perceived academic competence, appropriateness of instruction (reading), and learning environment (cooperation, student responsibility) have reasonably high positive correlations with engagement. These variables are all highly correlated with each other in grade 5, B-C.

When the global rating was partialled out of the fifth grade data, the correlation between orientation toward affect and percent engaged time changed from .01 to $-.55$ for reading and from .26 to $-.38$ for mathematics. When the global rating was partialled from academic competence, the correlation dropped from .60 to .22 for reading and from .71 to .18 for mathematics. The global rating appears to have a different multivariate relationship to engaged time and academic competence than to engaged time and orientation toward affect.

In order to better understand these relationships, we utilized a conjoint analysis of perceived academic competence, orientation toward affect and engagement. Figure 5-2 shows this multivariate relationship for the B-C period. The horizontal axis is perceived academic competence, unresidualized (G2UN). A vertical axis is orientation toward affect, unresidualized (G3UN). A letter represents each class. The percent of engaged time averaged across reading and mathematics in each class is marked near the letter in the plot. The mean percentage engagement for each quadrant of the plot is also provided. Classes above the standardized mean of 50 on both G2UN and G3UN had a mean engagement rate of 76.7%; classes below the mean on both G2UN & G3UN had a mean engagement rate of 69.7%, etc. Looking at the lower left quadrant, we note that teacher N had a score of approximately 35 on G2UN (academic competence) and a score of approximately 35 on G3UN (orientation toward affect). Both of these scores are well below the average mean scaled scores of 50. Class N had a mean engaged time of 73 percent. This data display shows the conjoint relationship of two generic classifications of teaching, or teaching environments, that have had a long history of value conflict. What should be emphasized in teaching: academics, affective support, or a blend or mixture of both?

Figure 5-2

Class Engagement Levels Plotted According to Perceived
Academic Competence and Orientation Toward Affect
(Grade 5, B-C Period)



5-39

G2UN Perceived Academic Competence, Unresidualized

Letter entries identify each class.

Numbers are average engagement across reading and mathematics.

An examination of the plot in Figure 5-2 indicates that relatively high scores on academic competence (G2UN) are positively associated with higher percents of engaged time. Scores on orientation toward affect (G3UN) show a small negative association with engaged time. The results are interactive, however. With high academic competence, emphasis on affect is associated with slightly lower student engagement. However, the major differences or contrasts appear to be between low G2UN in combination with high G3UN (upper left quadrant) and the other three quadrants. In classes with relatively low perceived academic competence and relatively high orientation toward affect, engagement is low.⁵

The interactive effect of orientation toward affect is highlighted in the residualized affect variable. That part of the affect rating which is independent of G1, that is, the residualized variable, identifies classes which are distinguished by having particularly high or particularly low affect scores relative to ratings on other (academic) variables. It intensifies these extremes (classes Q and E are high and low extremes on residualized affect). The high negative correlation between residual affect and engagement can be interpreted as indicating that when affect is emphasized to the exclusion of academic instruction, engagement is low.

In the plot of B-C data, only class E appears in the upper left quadrant (low academic competence, high orientation toward affect). In the A-B period, classes D, M, and T were also located in this quadrant near the boundary lines. They had engagement rates ranging from 53% to 66%. The B-C plot emphasizes the key role of class E in determining the negative

⁵It is important to remind the reader that the measures of academic competence and orientation toward affect are in a relative metric. They do not indicate what quantity of the characteristic being measured is associated with a given level of engagement.

relationship between orientation toward affect and engagement. With a small number of classes, it could be that the relationship occurs only by chance. It could also be that this cell is generally infrequently represented in a population; but when it occurs, engagement is low. Some support for the existence of the negative relationship can be found in the trend toward a negative relationship between affective emphasis and engagement in grade two. (See Table 5-12). It is interesting to speculate that warmth and attention to feelings may be hard to foster without a groundwork of academic accomplishment.

In attempting to further understand the complex relationships described above, we went to the files of descriptions of the BTES classes. Each field worker was asked to describe the classes and learning environments that they had closely observed over the year. The following edited resumes from the files describe four classes, one from each quadrant. Each of the classes described appeared in the same quadrant during the A-B and B-C periods.

Class K was high on both perceived academic competence and orientation toward affect during the B-C as well as the A-B period and had a moderately high average percent engagement (78%). Class K was described as follows:

"Warm, high academic direction of pupils, encourages pupils in higher cognitive tasks, shows intellectual curiosity, allows for individual pupil differences."

Class S (low academic competence and low orientation toward affect and moderately low average percent engagement, 73%):

"Disorganized approach to teaching, rarely attends to students' comments and questions, teaching is minimal and ineffective, shows little attention to pupil feelings."

Class Q (high perceived academic competence, relatively low orientation toward affect and very high (83%) average engaged time):

"Highly traditional teacher, all students work hard, orientation of pupils is almost totally subject matter, pupils appear to be highly motivated."

Class E (low perceived academic competence, relatively high orientation toward affect and low percent engaged time, 49%):

"Teacher appears to be 'burned out,' did not utilize academic materials the teacher had made 'previously, expressed awareness of lack of teaching, teaching was unorganized, frequently lost patience with students, knew a lot of personal history of each student."

Summary

This chapter reports on the relationships between teaching process measures and student engagement rate. Student engagement rate is the percentage of instructional time in reading or mathematics when the student appears to be paying attention to the academic task. The data in the previous chapter support the conclusion that students who pay attention more learn more. This chapter examines patterns of association between instructional variables and student engagement, in order to make inferences about what teachers do to foster high levels of student engagement.

One set of analyses looked at the relationships between student engagement and interactive teaching behaviors representing the functions of presentation, monitoring, and feedback. Descriptive statistics indicated that, on the average, students in our sample spent about two-thirds of their time working independently, with no instructional interaction taking place (e.g., doing workbook pages, reading silently, writing a report). Some students spent much more or much less time than this working independently (the range was from around 40% to around 90%). In general, there was a strong, positive relationship between the amount of substantive interactive instruction the student received and student engagement (correlations of around .45 in all four quadrants). This was probably the major finding in the chapter. Students who receive more interactive instruction tend to have higher rates of engagement than students who receive less interactive instruction. The following more specific findings were obtained:

1. Academic monitoring showed the most consistent relationship with engagement. In all four quadrants and both time periods, the correlation was significant ($\geq .18$). The more frequently a teacher asked questions of the students, the higher the rate of engagement.

2. Academic feedback was also generally positively related to engagement. The more frequently a student received feedback on the correctness of his answers, the higher his engagement. Academic feedback occurred more often in reading than in mathematics, largely because of oral reading. Academic feedback tended to be especially strongly related to engagement in reading.
3. Two presentation variables, structure/direct and explanation-planned, were associated with engagement in mathematics but not in reading. Explanation-planned occurred more frequently in mathematics than in reading. In mathematics, students were engaged more often when they received more explanation of concepts and procedures.
4. Task engagement feedback had a consistent, negative association with engagement. Task engagement feedback turned out to be primarily negative feedback, that is, it occurred at moments when students were off-task (statements like "Johnny get back to work.") Task engagement feedback was given more often to students who had generally lower rates of engagement. It was given more often on days when the student was more frequently off-task. Although task engagement feedback presumably limits the amount of off-task behavior, high percentages of task engagement feedback appear to signify a general lack of control.
5. Most interactive instruction took place in a group setting, with a relatively small part coming from one-to-one tutorial during seatwork. Analyses of setting indicated that engagement rates were higher when students spent more time in a group setting. However, amount of substantive interaction was a more potent predictor than amount of groupwork. When group time was more nearly filled with substantive

interaction, as opposed to organizational tasks or waiting for others, then engagement rates were higher during groupwork and during seatwork. When students received more contact with an instructor during seatwork, then engagement rates were higher in seatwork. Engagement rates were especially low when students spent over two-thirds of their time in seatwork and received little attention.

A second set of analyses looked at relationships between general classroom context and student engagement. One general variable was a global rating of the classroom made in response to the question "Would you like your child to be in this class?" The four other variables were appropriateness of instruction to student needs, perceived academic competence, orientation toward affect, and learning environment (cooperation, student responsibility). All variables were ratings. The following results were obtained:

1. General variables were not significantly related to engagement in grade 2.
2. In grade 5, observers gave more positive global ratings to classes with higher rates of engagement. Apparently, their image of a well-run grade 5 class included high student involvement in academic tasks; they expected to see students working hard.
3. In a follow-up of the global rating in grade 5, higher levels of student engagement were found to be associated with higher ratings of perceived academic competence and with lower ratings of orientation toward affect. These two effects were interactive. Particularly low engagement rates were observed in classes with below average ratings of academic competence and above average ratings of orientation toward

affect. One might conclude that to increase student engagement it is important for the teacher to have a clear focus on cognitive tasks and to provide competent academic instruction. When this groundwork is present, it is possible to attend to student feelings and create a warm environment and still maintain high engagement. When the teacher's primary focus is on affect, to the exclusion of academic skill teaching, then engagement is low.

⁶The BTES study is concerned primarily with academic achievement outcomes, and intermediate outcomes like engagement rate which are related to academic achievement. A teacher who focuses on the affective domain undoubtedly values other outcomes, such as self-concept, attitude toward school, artistic and social development. We have only limited information on such outcomes and thus can't report on the possible benefits of an affective orientation. It does seem clear from the data, however, that a choice is involved. When a teacher places primary emphasis on the affective domain, then both teacher and students pay less attention to academic instruction.

CHAPTER 6: TEACHING PROCESS VARIABLES AND STUDENT ERROR RATE

Another major component of ALT is student error rate. Students who show higher gains in achievement spend more time working on tasks that they understand well enough to complete correctly, i.e. where they make only chance, careless errors. This means they spend less time over the year working on tasks that they only partially understand and where they make a moderate number of errors. Students with higher achievement gains also spend less time on tasks that they do not understand at all and where they make many errors.

We hypothesize that student error rate can be influenced by instructional variables of two types. The planning functions of diagnosis and prescription should be related to student error rate. When teachers know more about their students' academic strengths and weaknesses and plan instruction that is responsive to student levels of performance, then this should eliminate the occurrence of high error rate tasks and promote more low error rate tasks. Also, interactive teaching behaviors may play a role. Students may make fewer errors when teachers take time to explain concepts, monitor student progress, and/or provide more feedback. Analyses were carried out to examine the effect of both these types of variables.

Before moving on to a discussion of instructional variables and of the complex relationships that emerged in these analyses, it is useful to consider again the nature of the error rate construct, especially low error rate. The variable that we use is an average score over five or fifteen weeks. Probably, the number of errors a student makes will shift up and down several times during that interval. This is because of the

periodicity inherent in error rate. Learning is a process of moving from not knowing to knowing. Most likely, when new material is introduced, the student will not understand completely and will make some errors. Guided practice and/or explanation help the student understand, and s/he comes to make fewer errors. Eventually, the student will perform correctly, although probably with some effort; and learning will become well established. This stage could be viewed as one of consolidation. At some point, the student knows the material so well that further practice is of minimal value (except for review), and it is time to move on to something new. BTES results suggest that for learning of basic skills in the elementary grades, the stage of successful practice is particularly important, so that concepts and procedures are thoroughly mastered. Apparently some teachers do not devote sufficient time to this stage. Such conclusions are inferred from variables that summarize data over many weeks. The low error rate variable, for instance, estimates the proportion of time during the B-C period (or A-B period) that a student spent on tasks at a low error rate. If a teacher (or curriculum) moves students through the cycle quickly, students will spend relatively less time on successful practice and relatively more time working on new, medium error rate material. In this case, the overall proportion of low error rate will be relatively low. Similarly, if a teacher assigns students difficult and challenging material, either by design or unintentionally, then the overall proportion of low error rate will be low. The data in Chapter 4 indicate that, on the average, students spent about half their time working on tasks where they had a low error rate (slightly less than half in grade 5, especially in math). Achievement gains were higher when students spent time working at a low error rate, i.e., more than half the time. Achievement gains were also higher when students were not given tasks that were extremely difficult (high error rate). The analyses reported

in this chapter are intended to identify instructional variables that are related to student error rate.

DIAGNOSIS AND PRESCRIPTION AND STUDENT ERROR RATE

The Diagnosis and Prescription Variables

Measures of diagnosis and prescription were derived from a variety of sources -- interviews with the teachers about planning, weekly interviews with the teachers, teacher tests and field worker ratings. A large set of measures of diagnosis, prescription, and general characteristics of the classroom was assembled, as noted in the previous chapter. A detailed process of variable reduction and definition was carried out, through which we selected for relational analysis three diagnosis variables and three prescription variables. A description of this process of defining variables can be found in Technical Note VI-1b (Filby and Cahen, 1977). The diagnosis and prescription variables are listed in Table 6-. All six variables are overall, yearly estimates for a particular dimension and thus were identical in A-B and B-C analyses.

Diagnosis. Diagnosis refers to the teacher's ability to recognize the strengths and weaknesses of individual students. The first variable in the diagnosis domain, item prediction, is perhaps the most direct measure of the teacher's knowledge about individual students' academic skills. The item prediction variable is a measure of the teacher's ability to predict how each target student will perform on a representative set of items from the BTES battery. Three different item prediction scores were created (see Chapter 3). This variable is a composite of the three scores. A high score on the item prediction measure means that a teacher is relatively more accurate in predicting test performance

Table 6-1

Definitions of Diagnosis and Prescription Variables

Variable		Method of Data Collection	Description
D1	Item Prediction	Teachers predicted student test performance (correct/incorrect) on 4 or 5 representative items from the BTES battery in reading and in math in the fall and in the spring. Teacher prediction was compared with actual performance.	Estimates teacher knowledge of the specific skills mastered by individual students and the level of proficiency of students in important areas of reading and mathematics.
D2	Knowledge of Subject Matter	Teachers completed a content coding test during training for log-keeping. The test involved labeling the content covered on typical student work sheets.	Estimates teacher knowledge of the terminology used to describe different concepts and skills in reading and mathematics (such as "consonant blends," "main idea," "short-form regrouping").
D3	Differentiated Perceptions	A field worker rating based on interviews and observation. (7-point scale; average over the six target students)	Describes the extent to which the teacher is aware of individual differences and needs of target students.
P1	Goal Orientation	A combination of 2 field-worker ratings based on interviews and observation.	These correlated ratings describe the extent to which the instructional program plans activities toward specific goals and tests to see that goals are reached: Goal Directedness Content Goal Assessment
P2	Program Change Based on Need	Each week teachers were asked to report any changes in the target students or in their instructional program. The number of statements of different types were added up over the year.	Describes the frequency with which teachers reported changing the instructional program because they noticed that the student(s) were progressing quickly or needed extra help.
P3	Flexibility in Use of Curriculum	A field-worker rating based on interviews and observation. (7-point scale).	Estimates the extent to which the teacher changes materials or ways of using materials based on student responses in order to adjust to ongoing instructional needs.

than the other teachers in the sample.

The second diagnosis variable is knowledge of subject matter. This is our measure of "aptitude." It comes from the teachers' performance on a content coding test. Knowledge of subject matter is a less direct measure of diagnosis than is item prediction. It reflects the teacher's grasp of the concepts and terminology of reading and mathematics pedagogy. Such knowledge presumably provides a framework for diagnosis.

The third diagnosis variable is an interview-based rating of differentiated perceptions. The score is residualized on the global rating, G1, "Would you like your (hypothetical) child to be in this class?" (Because all ratings were highly correlated with this "halo" rating, residual scores were created which partial out G1; see the section on general variables in Chapter 5 for details). This rating reflects the degree to which the teacher is aware of and comments on individual characteristics and needs of different target students. It is not necessarily cognitive in focus and says nothing about the accuracy of the teacher's differentiations. Presumably, like knowledge of subject matter, perception of individual students provides a precondition for diagnosis.

Prescription. The prescription function is perhaps the most complex to assess. It refers to the teacher's decisions about what to do in the classroom. One way to study prescription would be to look at the content of the decisions that are made. This would lead to studies of the efficacy of different curricula or organizational factors. For instance, a decision to have three reading groups or to use a programmed math kit represents prescription. Studying such factors was not considered to be the focus of BTES.

Another way to look at prescription is to focus on the decision-making

process itself, how decisions are made. This aspect of prescription is what is labeled prescription for these analyses. Three variables were selected to describe specific aspects of the way in which prescription decisions are made. These variables are listed in Table 6.1. All prescription variables are defined over the A-C period.

Variable P1 is a rating of goal orientation. This variable is a combination of two interview-based ratings, goal directedness and content goal assessment. Goal directedness rates the extent to which the instructional program identifies specific goals for learning activities, that is, the extent to which goals guide the selection of daily tasks. Content goal assessment rates the extent to which the program includes testing to see if goals are reached and to specify the next instructional steps. These two ratings were highly correlated, and were combined to indicate the extent to which the program is goal oriented. In a highly goal oriented program, instruction is an orderly progression toward clearly stated goals.

The second prescription variable is the weekly interview variable, program change based on need. On each observation day, the teacher was asked if any changes had occurred in the target students or in their programs. One type of response was to report a change in a student's program because of something that had been observed about the student's performance. For instance, the teacher might have noticed that a student was having trouble with some concept and assigned special homework. The score used for these analyses was the average across target students of the number of program-change-need statements made throughout the year. This variable provides a relatively objective measure of the extent to which teachers decide to make changes based on ongoing observation of student performance.

The third prescription variable is a rating of flexibility of use of curriculum materials. It differentiates between teachers who strictly follow the procedures and sequence of instruction contained in a set of curriculum materials and teachers who modify materials and procedures on an ongoing basis depending on student responses. It reflects the extent to which decisions are made by choosing a curriculum and following its mandates as opposed to using a more complex, ongoing, adaptive decision framework.

In addition to these specific prescription variables, the general variable called appropriateness was also used in these analyses. Appropriateness, described in Chapter 5, represents the integration of diagnosis and prescription on an ongoing basis.

Descriptive Statistics

Means, standard deviations and reliabilities of diagnosis variables are given in Table 6-2. These statistics for prescription variables are given in Table 6-3. All six variables were standardized to a mean of 50 and a standard deviation of 10, as were the general variables introduced in Chapter 5. Such a transformation does not alter the underlying distribution of the variable or the relationship of the variable to other variables. Standardization does make it easier to compare a teacher's scores on different measures.

The reliability of an individual variable in a particular quadrant sometimes drops below the level one would like. Reliability is lower on D3 and P3 because they are residual score variables and thus contain a higher proportion of error. This trade-off was considered acceptable because of the greater conceptual clarity of a variable with "halo" removed. Also, we decided to use the same variables for analysis in all four

Table 6-2
Means, Standard Deviations, and
Reliabilities of the Diagnosis Variables

Variable	Mean	Standard Deviation	Reliability			
			Grade 2 Rdg.	Math	Grade 5 Rdg.	Math
D1 Item Prediction	50	10	.76	.77	.76	.62
D2 Knowledge of Subject Matter	50	10	+	+	+	+
D3 Differentiated Perceptions*	50	10	.61	.71	.27	.36

+ Unavailable

* This variable is a residual score derived by regressing the original rating on G1, the global rating.

Reliability of D1 is the reliability of a composite of three measures.
Reliability of D3 is inter-rater reliability, based on two ratings of each class.

Note: All three diagnosis variables are defined A-C; therefore, the variables are identical in A-B and B-C analyses.

Table 6-3
Means, Standard Deviations, and
Reliabilities of the Prescription Variables

Variable	Mean	Standard Deviation	Reliability			
			Grade 2 Rdg.	Math	Grade 5 Rdg.	Math
P1 Goal Orientation*	50	10	.64	.37	.69	.75
P2 Program Change Based on Need	50	10	.73	.82	.75	.89
P3 Flexibility in Use of Curriculum*	50	10	.58	.32	.41	.45

* These variables are residual scores derived by regressing each of the original ratings on G1, the global rating.

Reliability of P1 and P3 is inter-rater reliability, based on two ratings of each class. Reliability of P2 is inter-coder reliability based on two codings of weekly interview protocols.

Note: All three prescription variables are defined A-C; therefore, the variables are identical in A-B and B-C analyses.

quadrants. If estimated reliability dropped in one or two quadrants, we decided to keep that in mind for interpretation but not to eliminate the variable from some analyses.

Intercorrelations of the six diagnosis and prescription variables are given in Tables 6-4 and 6-5 for grade 2 and grade 5 respectively. D1, item prediction, and D2, knowledge of subject matter, are always positively correlated, although the magnitude of the correlation is not striking. Differentiated perception, D3, is strongly correlated with item prediction in grade 2. In grade 5, this relationship shifts for some reason to a low negative one. The program-change-need variable (P2) is generally negatively correlated with the other two prescription variables and with the diagnosis variables in grade 2. This is counterintuitive and suggests that the variable may be measuring something more complex than the label implies.

Correlations with Low Error Rate and High Error Rate

To investigate the relationships between planning variables and student error rate, we began with a set of commonality analyses using regression techniques to determine the relative impact of the set of three diagnosis variables and the set of three prescription variables. These analyses and significant unique regression weights are reported in Technical Notes V-1b and V-2b (Filby and Cahen, 1977, 1978). We also looked at correlations, and these results are repeated in this report, in Table 6-6. In addition, later analyses indicated the importance of the general variable appropriateness to student needs. Correlations of appropriateness with error rate are included in Table 6-6.

Two dependent measures were used. One was low error rate, the percentage of the time in reading or mathematics that the student was working on tasks where he made few errors, only those you would expect by chance.

Table 6-4
Intercorrelations of Diagnosis and Prescription Variables
(Grade 2)

	D1	D2	D3	P1	P2	P3
D1	<i>41</i>	31	58	06	11	05
D2	15	<i>34</i>	-08	-33	-02	-14
D3	40	27	<i>93</i>	40	-20	51
P1	-15	-03	31	<i>77</i>	-44	33
P2	-21	-37	-13	-01	<i>20</i>	-10
P3	25	23	55	15	-17	<i>41</i>

Entries below the diagonal are for reading. Entries above the diagonal are for mathematics. Entries in the diagonal, in italic type, show the correlation between reading and mathematics.

Sample size is 25 classes.

Note: All diagnosis and prescription variables are defined over the A-C period; therefore, the variables are identical in A-B and A-B analyses.

Table 6-5
Intercorrelations of Diagnosis and Prescription Variables
(Grade 5)

	D1	D2	D3	P1	P2	P3
D1	34	19	-28	02	06	-16
D2	20	28	06	38	11	15
D3	-24	06	88	07	12	47
P1	39	46	19	87	05	20
P2	19	01	-05	21	77	29
P3	-23	40	65	31	28	70

Entries below the diagonal are for reading. Entries above the diagonal are for mathematics. Entries in the diagonal, in italic type, show the correlation between reading and mathematics.

Sample size is 21 classes.

Note: All diagnosis and prescription variables are defined over the A-C period; therefore, the variables are identical in A-B and B-C analyses.



Table 6-6
Correlations of Diagnosis and Prescription Variables
with Low Error Rate and High Error Rate

Variable	Low Error Rate								High Error Rate							
	2R		2M		5R		5M		2R		2M		5R		5M	
D1 Item Prediction	18	<i>14</i>	-17	<i>29</i>	-23	<i>-17</i>	-10	<i>09</i>	-12	<i>-51*</i>	-06	<i>-20</i>	-29	<i>-26</i>	25	<i>-59*</i>
D2 Knowledge of Subject Matter	-50*	<i>-14</i>	20	<i>07</i>	-15	<i>07</i>	13	<i>05</i>	-16	<i>-06</i>	-37	<i>-08</i>	-22	<i>-06</i>	-03	<i>02</i>
D3 Differentiated Perceptions ¹	-26	<i>16</i>	-12	<i>19</i>	-02	<i>-05</i>	-04	<i>-11</i>	11	<i>-43</i>	02	<i>-16</i>	00	<i>14</i>	-01	<i>-12</i>
P1 Goal Orientation ¹	18	<i>41</i>	01	<i>-09</i>	-23	<i>-19</i>	10	<i>-44</i>	16	<i>-02</i>	-02	<i>05</i>	-40	<i>-12</i>	-20	<i>01</i>
P2 Program Change Based on Need	10	<i>-19</i>	-40	<i>-21</i>	41	<i>17</i>	35	<i>06</i>	11	<i>37</i>	39	<i>27</i>	15	<i>51*</i>	-07	<i>12</i>
P3 Flexibility of Curriculum ¹	-27	<i>27</i>	-01	<i>-19</i>	23	<i>19</i>	25	<i>-05</i>	30	<i>-22</i>	-04	<i>07</i>	-03	<i>28</i>	-40	<i>07</i>
A Appropriateness ¹	14	<i>22</i>	-03	<i>-02</i>	-16	<i>-27</i>	-15	<i>-46*</i>	-19	<i>-12</i>	-01	<i>-15</i>	-53*	<i>-11</i>	-40	<i>-21</i>

The numbers in italics are the correlations obtained in the A-B period.

Sample size is 25 classes in grade 2 and 21 classes in grade 5. Correlations reach significance at or beyond the .05 level when $r > .45$. (The level of significance will differ somewhat as a function of sample size and slightly due to estimation of missing data. Correlations $> .45$ are marked with an asterisk ().

¹These rating variables are residual scores with the global rating partialled out.

6-13

The other was high error rate, the percentage of the time that the student was working on tasks that s/he was unable to do at all correctly. The reader should keep in mind that student time in this study was generally divided between low error rate and medium error rate. The percentage of the time when high error rate occurred was relatively small overall, but diagnostic of learning (or rather, lack of learning). Because low error rate and medium error rate were so highly correlated ($\geq .95$), only one of the two variables needed to be used in analysis. We used low error rate because it was the positive predictor. It can be interpreted as meaning more low and less medium.

Because the diagnosis and prescription variables are teacher or class-level variables, all analyses were carried out at the class level. Class averages on low error rate and high error rate were calculated and used as dependent measures. (This has the added, fortunate effect of making the dependent measure more reliable).

Correlations were calculated separately for the A-B period and the B-C period. All planning variables are defined year long, but period-specific error rate estimates were used. Correlations between planning variables and student error rate did not yield many strong or consistent relationships. In particular, relationships varied across time periods. For high error rate, there were two relatively strong and consistent relationships in the A-B period, but these relationships did not appear as strongly in the B-C period. One variable was consistently negatively related to high error rate, but was especially strong in the B-C period.

During the A-B period, D1, item prediction, was negatively associated with high error rate. In classes with more high error rate, teachers were less accurate in prediction of student performance on BTES achievement test

items. One inference from this result is that teachers who are good diagnosticians avoid giving students tasks that are extremely hard. This relationship reached statistical significance in two of the four quadrants and was consistent across quadrants in the A-B period. This trend was much less strong in the B-C period and reversed direction in grade 5 math. Since D1 is the same variable, it must be the case that high error rate has changed. Indeed, data in Chapter 3 shows that high error rate, a relatively infrequent variable, is relatively unstable. The correlation at the student level between A-B and B-C values of high error rate is around .25 in each grade. An examination of the plots indicated that classes which were extremes in the A-B period sometimes shifted position dramatically in the B-C period. One explanation for this changing relationship is that the causes of high error rate shift during the year. Early in the year, lack of knowledge of the students may produce inappropriate instruction. Teachers who use a diagnostic-prescriptive approach may make a special effort early in the year to get to know what individual students can do. Other teachers may begin instruction based on a general sense of student performance levels and learn more slowly by trial and error what levels are appropriate for different children. Later in the year, curriculum variables may be more closely associated. Knowledge of subject matter tends to be somewhat more strongly negatively associated with high error rate in the B-C period. Knowing something about the sequential progression of curriculum may help the teacher avoid introducing material that is too hard. The reader should note that while these explanations have some plausibility the relationships in the data are not strong.

The other significant correlation in the A-B period was a positive correlation between P2, program change based on need, and high error rate.

As with item prediction, the shift in high error rate in the B-C period reduced the size of the correlations so that none of them achieved conventional levels of significance. The trend was generally in the same direction in both periods. Teachers made more program changes based on student need in classes where students were more often unable to do their assignments. Apparently, program change based on need tends to be a response to extreme difficulty.

Some insight into this relationship can be provided by considering the correlations between P2 and other instructional variables. As shown in Tables 6-4 and 6-5, program change based on need tended to have low positive or even negative correlations with other diagnosis and prescription measures. Appendix 7-B gives correlations with general and interactive variables. Program change based on need tends to be positively correlated with observed frequency of explanation in response to need. This provides some validation for the teacher-report measure. It also tends to be positively correlated with frequency of interactive instruction (groupwork) but negatively correlated with measures of academic competence (such as G2 in grade 2 and G1 in grade 5). One can interpret this pattern as indicating that some teachers use group instruction but not in a highly competent manner. Picking up on the explanation offered for item prediction, these teachers may start off the year with group instruction and without a very thorough understanding of student abilities to guide assignment of students to groups. In this situation, some students get assignments they can't do. The teacher then responds to this situation by trying to give extra explanation. If the situation is bad enough, the teacher will make some kind of change, such as changing group assignment or giving extra work. The general tone, however, is one of little pre-planning, group instruction, and teacher response only

when pressed. The danger in this approach is that the response will be too little, too late.

One variable had a consistent negative relationship with high error rate across both time periods. This variable was appropriateness of instruction to student needs. The negative relationship was particularly strong in grade 5 during the B-C period ($r = -.53$ in grade 5 reading B-C). The instructional program was judged less appropriate when students were given tasks they could not do. Appropriateness to student needs is a field worker rating based on interviews and extensive observation in the classroom. It estimates the extent to which the teacher uses diagnosis to plan instruction which meets the needs of individual students. It stresses the ongoing use of diagnosis throughout the year to adapt instruction to the needs of the students. A high rating of appropriateness would be given if the teacher created small groups and was flexible in the assignment of children to groups. A high rating would be given if the teacher used an "individualized" program where students worked on different assignments at different rates, as long as the assignments did appear appropriate for each student. Teachers who manage instruction so that diagnosis continually guides prescription are less likely to assign tasks that are too hard. This type of system was particularly important for eliminating high error rate in the second half of the year.

The appropriateness variable synthesizes the interpretations offered for the last two variables. Use of a diagnostic-prescriptive approach implies diagnosis early in the year and a fast start in terms of matching students to appropriate content. This reduces the occurrence of high error rate in the A-B period. It also implies continued monitoring of student performance and flexibility in the system so that students are kept on track. This reduces

the occurrence of high error rate in the B-C period. Program change based on need, as it happened in our data, tended to reflect changes that were "too little, too late," and not based on a diagnostic-prescriptive framework. Flexibility within that framework, to keep students working on appropriate material, is desirable.

Appropriateness of instruction to student needs was also related to low error rate. Appropriateness tended to be negatively associated with low error rate in grade 5, especially in the A-B period. One of the correlations was statistically significant ($r = .46$ in grade 5 math, A-B). All four grade 5 correlations were negative. In classes where instruction was judged more appropriate students spent relatively more time at a moderate error rate and less time at a low error rate.

One possible interpretation that should be considered is that raters were influenced by the ALT model. During the time that data were collected, the prediction was that medium error rate would reflect maximum learning, and the directions for the appropriateness rating scale specifically mention considering error rate. However, the difference between correlations in grade 2 and grade 5 suggests that this rating was a multi-faceted judgment that was influenced by model predictions only moderately if at all.

Apparently, medium error rate was considered more appropriate in grade 5 than in grade 2. Some insight into this finding can be gained by considering the intercorrelations among instructional variables in grade 5.¹ Especially in grade 5 reading, appropriateness was high correlated with academic focus and with student responsibility for academic achievement. Some teachers in grade 5 stressed academics and expected students to be able to cope with

¹ A matrix of intercorrelations can be found in Appendix 7-B.

challenging tasks. Those teachers who strongly emphasized academics and who individualized instruction appropriately tended to give more medium error rate tasks.

Among the correlations of other planning variables with low error rate, two relationships in the B-C period are worth noting briefly.

Knowledge of subject matter, D2, had a statistically significant negative correlation with low error rate in grade 2 reading in the B-C period ($r = -.50$). This relationship did not appear in the other quadrants and was less strong in the A-B period. In the absence of a consistent trend, it seems highly possible that the significant correlation occurred by chance. If the relationship is real, it could be interpreted as indicating that teachers who know more of the concepts and terminology of skill instruction in reading assign more medium error rate tasks and fewer low error rate tasks. (Perhaps this effect is mediated by the tendency to teach a variety of skills).

The other relationship with low error rate in the B-C data was a trend toward a positive association in grade 5 between responsiveness/flexibility and low error rate. The strongest correlation was between P2, program change based on need, and low error rate in grade 5 reading ($r = .41$). These two variables, P2 and low error rate, were also positively correlated ($r = .35$) in grade 5 math. In addition, P3, flexibility of curriculum was positively associated with low error rate in these two quadrants. P2 and P3 are intercorrelated about .29 in grade 5. When combined in a multiple regression, both contribute independently and the multiple regression is statistically significant. One of these variables, P3, is a rating. The other is a count of interview statements. Both reflect the teacher's tendency to alter the instructional program based on how students respond. In grade 5, a

tendency toward responsiveness/flexibility seems to be associated with more low error rate tasks.

This pattern of results needs to be considered in relationship to the appropriateness variable. P2, P3, and A all contain some element of responsiveness/flexibility but they have very different correlations with low error rate in grade 5. What does this mean? Appropriateness tends to be associated with an emphasis on academics, perceived academic competence and, in grade 5, an individualized program. The whole instructional system is organized so that individual students move at their own rate through a sequential curriculum. At least in those BTES classes, students were expected to work on moderately difficult (but not high error rate) tasks. Flexibility is built into a system which is highly goal oriented. Program-change-need, as we have seen, tends to be associated with groupwork and with teachers not wanting things to be too hard, although the problems of grouping students sometimes produced that. Flexibility of use of curriculum was correlated with orientation to affect. These teachers were concerned with what the students liked or were interested in. They would do things like use the most interesting stories in the reader or extend a lesson if the students were enjoying the discussion. Presumably, these teachers used easier material because they wanted students to like what they were doing. (Data in Chapter 9 suggests they were right!)

INTERACTIVE TEACHING BEHAVIORS AND STUDENT ERROR RATE

The second set of analyses of teaching process measures and student error rate focused on interactive teaching behaviors and their relationship to error rate. The interactive teaching behaviors represent the functions of presentation, monitoring and feedback. Presumably, students will make

fewer errors on daily tasks if concepts are explained, progress is monitored (so the teacher knows that the student is making errors), and feedback is given to correct errors.

The six interactive teaching behaviors were introduced in Chapter 5. The behaviors were defined in Table 5-1. Exactly these same behaviors were used in this set of analyses. The two dependent measures were low error rate and high error rate. Both the interactive teaching behaviors and error rate were derived from observation of individual students. Since both sets of variables were measured at the student level, analyses were carried out at the student level.

Again the analysis strategy involved commonality analysis by function to determine general effects at the function level. Results of these analyses and the unique contribution of individual behaviors in multiple regression were reported in Technical Note V-2b (Filby and Cahen, 1978). Analyses of interactive teaching behaviors and error rate were carried out only for data from the B-C period. Correlations of individual behaviors with low error rate and high error rate are reported in Table 6-7.

Three relatively strong trends can be identified in the data. The first is for the variable explanation-need. Explanation-need is consistently negatively associated with low error rate. The relationship is statistically significant in both mathematics quadrants and negative in all four quadrants. Students with less time coded as easy receive more explanation based on need. Explanation-need is consistently positively associated with high error rate. The relationship is statistically significant in grade 2 mathematics and grade 5 reading. Students with more time coded as hard receive more explanation based on need.

One interpretation is that these results validate the definition of

Table 6-7

Correlations of Interactive Teaching Behaviors
with Low Error Rate and High Error Rate
(in the B-C Period)

Variable	Low Error Rate				High Error Rate			
	2R	2M	5R	5M	2R	2M	5R	5M
Explanation-Need	-10	-35*	-06	-33*	05	35*	29*	12
Explanation-Planned	-06	-13	04	-14	01	07	28*	04
Structure, Direct	06	17	29*	11	-00	-21*	23*	16
Academic Monitoring	-12	-17	09	16	-06	-01	28*	16
Academic Feedback	08	-06	-10	07	-15	-06	-03	07
Task Engagement Feedback	12	-31*	15	-25*	-02	19*	45*	27*

Sample size is 139 students in grade 2 and 122 students in grade 5.
Correlations reach significance at the .05 level when $r > .17$. Significant correlations are starred.

the variable. Students receive explanation-need when they need it, when they are still making mistakes and do not yet fully understand the material. One would have hoped, of course, that explanation had been given in a timely enough manner to reduce confusion overall and increase the percentage of the time coded as low error rate. Such a situation would produce a positive correlation between explanation-need and low error rate and a negative correlation with high error rate. But this is not what happened. The interactive teaching behavior explanation-need is a response to an observed student condition. It attempts to remedy misunderstanding. It does not, relative to other patterns of instruction, produce superior understanding. Apparently the response is too late, a reaction to a situation that could be prevented from occurring in the first place.

The second variable related to difficulty level is structure/direct. Structure/direct is generally positively related to low error rate.

Tasks are more often coded easy for students who more often receive task structuring or directions from the teacher. This relationship is less strong than relationships between explanation-need and error rate. The average of the four correlations is only 16, not statistically significant at the .05 level; but the trend is consistent across quadrants. Possibly structuring helps the child recognize the task and recall the knowledge needed to perform the task. Providing directions may help particularly in avoiding errors due to task format. The student may have a basic understanding of the content but make errors on a task because he does not follow specific task requirements. This would probably result in medium error rate. Giving directions would reduce this type of misunderstanding.

With high error rate as a dependent measure, structure/direct has a contradictory set of correlations, sometimes negative, sometimes positive.

It may be, following the line of previous discussion, that unskilled use of groupwork is associated with more high error rate in grade 5 (see the discussion about program change based on need). That would fit the general pattern of correlations in grade 5. This pattern does not seem to exist in grade 2.

In the feedback domain, task engagement feedback is the specific behavior which sometimes produces significant relationships. With low error rate, the pattern of relationship is mixed. In math, task engagement feedback is negatively correlated with low error rate. This direction for the relationship would seem to make the most sense. From the analysis of task engagement feedback and student engagement, we know that task engagement feedback is primarily negative and tends to occur in response to off-task behavior. (See page 5-13 and the discussion following). We also know from Chapter 4 that there is a slight tendency for less low error rate to be associated with less engagement. Therefore, it would seem most likely that task engagement feedback would occur more often for students with fewer low error rate tasks. At the extreme of the scale, this does seem to be the case. In three of the four quadrants, high error rate is positively associated with task engagement feedback. Students who make many errors on some tasks receive more task engagement feedback. This pattern of results is compatible with the results for engagement. Notice the particularly strong association in grade 5. One interpretation, putting together several results, is that some students in grade 5 are put in a group setting but given tasks they cannot do. These students then misbehave and are reprimanded by the teacher, who notices their behavior more in the group setting. A number of field worker comments provide concrete examples of this unfortunate pattern.

Summary

This chapter examined the relationships between teaching process measures and student error rate. Student error rate measures the difficulty of daily tasks for each individual student. As the student worked, the observer evaluated error rate according to three broad categories -- "low," the student performed correctly except for some careless errors, "medium," the student knew enough to work on the task but did not always answer correctly, and "high," the student did not know how to do the task. Data were averaged across daily tasks and across observation days to get the percentage of the time that each student was working at low, medium and high error rates. We found that, on the average, time was divided between low and medium error rates. High error rate was relatively infrequent. Analyses in Chapter 4 indicated that students show higher achievement gains when they spend more time performing at a low error rate rather than medium. Analyses also showed that students have higher gains when they spend less time (effectively no time) at a high error rate. In this chapter we looked at relationships between teaching process measures and these two error rate variables -- percent low error rate and percent high error rate.

One set of analyses looked for relationships between student error rate and the teacher planning functions -- diagnosis and prescription. Measures of diagnosis assess the teacher's knowledge about the strengths and weaknesses of individual students. Prescription measures describe the way in which curriculum decisions are made. The question is whether diagnostic skill or manner of prescription help the teacher plan instruction so that students have lower error rates on daily tasks. In addition to specific diagnosis and prescription variables, a general planning variable was included in these analyses. This was the rating on "appropriateness of

the instructional program for individual student needs." This variable estimates the degree to which knowledge of individual students is used on an ongoing basis to prescribe apparently reasonable instructional programs. Presumably, appropriate instruction would involve more low error rate and less high error rate.

Correlations were reported for both the A-B and the B-C period. The same teaching process scores were used in both sets of correlations, but error rate was defined within period. The initial purpose was to look for replication of results. In fact, correlations differed quite a bit across time periods. This may indicate the instability of the error rate measures. It may also indicate that different factors are important at different stages during the year. In general, results were complex and this area is in need of further in-depth study. A number of possible explanations for specific results were offered in the body of the chapter. Results were as follows:

1. In the A-B period, the diagnosis variable of item prediction was consistently negatively associated with high error rate. The relationship reached significance in two of the four quadrants. In classes where students were sometimes unable to perform daily tasks, teachers were less accurate in predicting student performance on BTES achievement test items. This relationship was much less strong in the B-C data. Some teachers may make a particular effort in the fall to get to know student skills and thus avoid giving tasks that are too hard.
2. In the A-B period, the prescription variable of program change based on student need was consistently positively associated with high error rate. The relationship reached significance in one of the

four quadrants. It appeared that teachers tended to make program changes based on need when students were having great difficulty. This relationship was less strong in the B-C data, where it did not reach statistical significance. It was suggested that some teachers may start out in the fall with group instruction but without adequate diagnosis to guide group assignments. When things go wrong, they do something about it; but it may be too little, too late.

3. The observer rating of appropriateness of instruction to student needs was consistently negatively correlated with high error rate. The relationship was stronger in the B-C period, especially in grade 5, where it reached statistical significance. The instructional program was judged less appropriate when students were given tasks they could not do. Apparently, teachers who consistently use a diagnostic-prescriptive approach and try to match instruction to individual needs on an ongoing basis are successful in terms of avoiding tasks that are too hard.
4. Appropriateness also tended to be negatively correlated with low error rate, in grade 5, especially in the A-B period. One correlation was significant. Apparently, medium error rate was judged more appropriate. This could be influenced by the fact that the ALT model at the time these ratings were done hypothesized that medium error rate would produce learning. It appears, however, that medium error rate is considered more appropriate for grade 5 students than for grade 2 students. Moreover, those fifth grade

teachers who emphasize academics and expect high levels of work orientation from the students are the ones who tend to individualize instruction and to assign challenging, medium error rate tasks.

5. There was a slight tendency in grade 5 for responsiveness/flexibility in prescription to be positively related to low error rate in the B-C period. More program changes based on the needs of students and more flexibility in use of curriculum were associated with a lower error rate for grade 5 students. Examination of the relationships among variables suggested that program-change need is related to use of group instruction. Flexibility in use of curriculum is associated with orientation toward affect. Apparently, those teachers in our sample who were concerned about student interests and wanted students to like school were more likely to adapt daily assignments and to give easier material.

A second set of analyses looked at the relationships between interactive teaching behaviors and student error rate. The interactive teaching behaviors represent the general functions of presentation, monitoring, and feedback. They are the same variables used in Chapter 5 to predict engagement rates. In this set of analyses, the question is what daily instructional events contribute to the accuracy of student performance on daily tasks. One would like to find teaching behaviors which reduce the number of student errors. The following results were obtained in analyses of the B-C data:

1. Explanation-need was consistently negatively associated with low error rate and consistently positively associated with high

error rate. Each relationship was significant in two of the four quadrants. Students who received more explanation specifically in response to needs or misunderstandings tended to have more tasks they could not do and fewer tasks they performed correctly. Another way of saying this is that students who had more need received more explanation in response to need.

2. Structure/direct was generally positively related to low error rate. The correlation was significant in one of the four quadrants. Students made fewer errors when they received directions from the teacher more often. This may be related to confusing response formats on worksheets.
3. Task engagement feedback was significantly positively associated with high error rate in three of the four quadrants. Tasks tended to be harder for students who received more task engagement feedback. A pattern of inappropriate grouping for instruction and subsequent off-task behavior was offered as explanation.

CHAPTER 7: TEACHING PROCESS VARIABLES AND ACADEMIC LEARNING TIME AS PREDICTORS OF STUDENT ACHIEVEMENT

From the last three chapters, we know that the facets of academic learning time are significant predictors of student achievement and we know something about those teaching process variables that are related to ALT. The assumption underlying such a two-stage analysis is that teaching process variables influence student achievement indirectly, by first influencing student ongoing learning behavior (ALT). In this chapter, we investigate this assumption by looking at the relationship between teaching process measures and student achievement and by considering how teaching processes and ALT together relate to achievement.

Correlations Between Teaching Process Variables and Student Achievement

As a first step, we looked at the relationship between each of the teaching process variables introduced in the last two chapters and residual achievement in reading and mathematics. Residual achievement scores were created on each subtest for each student over the A-C period. This was done by regressing each post achievement score at C on its pre achievement score at A to generate a predicted posttest score. The difference between the observed post achievement score and the predicted post achievement score for each student was that student's residual achievement score. Tables 7-1 and 7-2 identify the different subtests that were used in each quadrant and the pre achievement measure that was used in each case.

Correlations between teaching process variables and residual achievement are reported at the student level. For the planning and general variables, this means that a class-level variable is being correlated with

Table 7-1

Post Achievement, Pre Achievement, and
Time Variables Used in A-C Analyses
of Teaching Processes, ALT, and Achievement
(Grade 2)

Post Achievement (C Score)	Pre Achievement ¹ (A Score)	Allocated Time ²
Reading		
Decode Blends and Long Vowels	Decode Blends and Long Vowels	Decode Blends and Long Vowels
Decode Variant Consonants	Decode Variant Consonants	Decode Variant Consonants
Decode Complex Patterns	Decode Complex Patterns	Spelling
Word Structure - Meaningful Units	Word Structure - Meaningful Units	Word Structure - Meaningful Units
Syllables	Syllables	Syllables
Total Comprehension	Total Word Meaning	Total Comprehension
Total Reading	Total Reading	Total Reading
Mathematics		
Addition and Subtraction, no Regrouping	Addition and Subtraction, no Regrouping	Addition and Subtraction, no Regrouping
Addition and Subtraction, Speed Test	Addition and Subtraction, Speed Test	Addition and Subtraction, Speed Tests
Addition and Subtraction, With Regrouping	Place Value/ Numerals	Addition and Subtraction, with Regrouping - Algorithms
Computational Transfer	Computational Transfer	Computational Transfer
Place Value/Numerals	Place Value/Numerals	Place Value/Numerals
Word Problems	Word Problems	Word Problems
Money	Money	Money
Fractions	Computational Transfer	Fractions
Total Mathematics	Total Mathematics	Total Mathematics

¹This is the achievement measure used as a pretest. Residual achievement scores were created by regressing post on pre. The same pre measure was used in the multiple regressions reported in Table 7-10.

²This is the "matched" time category used in the multiple regressions reported in Table 7-10. Allocated time estimates come from teacher logs.

Table 7-2

Post Achievement, Pre Achievement, and
Time Variables Used in A-C Analyses
of Teaching Processes, ALT, and Achievement
(Grade 5)

Post Achievement (C Score)	Pre Achievement ¹ (A Score)	Allocated Time ²
Reading		
Total Word Meaning	Total Word Meaning	Total Word Meaning
Total Comprehension	Total Comprehension	Total Comprehension
Syllables	Syllables	Syllables
Total Reading	Total Reading	Total Reading
Mathematics		
Multiplication	Multiplication	Multiplication
Division	Multiplication	Division
Fractions	Multiplication	Fractions
Mult. Speed Test	Mult. Speed Test	Mult. Basic Facts
Total Geometry	Total Geometry	Total Geometry
Computational Transfer	Computational Transfer	Computational Transfer
Word Problems	Word Problems	Word Problems
Total Mathematics	Total Mathematics	Total Mathematics

¹This is the achievement measure used as a pretest. Residual achievement scores were created by regressing post on pre. The same pre measure was used in the multiple regressions reported in Table 7-11.

²This is the "matched" time category used in the multiple regressions reported in Table 7-11. Allocated time estimates come from teacher logs.

a student-level variable. For class-level variables, the same class/teacher score was entered for each student in the class. This handicaps the teaching variable in that it can never account for differences between students within a class. It also creates a situation where standard statistical models and interpretations are not entirely applicable. A problem exists in determining the appropriate degrees of freedom and level of significance associated with a correlation based upon one class-level variable and one student-level variable. Educational methodologists are now seriously considering issues of level of analysis and the separation of between-class and within-class relationships (Burstein, 1976 ; Cronbach, 1976). There is no consensus on how to deal with these issues. We looked at both student-level and class-level correlations and decided on a cut point for believing the existence of relationships. This procedure allows us to interpret correlations which in our view seemed to capture the major relationships regardless of level. The details of this procedure are described below. We report student-level correlations because of the consistent focus throughout the study on individual student learning and because we want to integrate class and student variables in a joint prediction of achievement.

The correlations between teaching process variables and student achievement are summarized in Tables 7-3 through 7-6. Tables 7-3 and 7-4 deal with planning and general variables in grades 2 and 5 respectively. Tables 7-5 and 7-6 deal with interactive teaching behaviors in grades 2 and 5. The format of all four tables is the same. The correlation with residual achievement over the A-C period in total reading and total mathematics is reported. In addition, a summary of significant correlations across subtests and the total test is provided. Each teaching process measure was correlated with residual achievement in each of the subtests listed in Tables 7-1 and 7-2.

Table 7-3

Summary of Correlations Between Planning or
General Variables and Residual Achievement

(Grade 2)

Variable	Grade 2 Reading			Grade 2 Mathematics		
	Correlation with A-C Residual Ach. (Total Test)	Number of Significant Correlations (out of 7) Positive Negative		Correlation with A-C Residual Ach. (Total Test)	Number of Significant Correlations (out of 10) Positive Negative	
D1 Item Prediction	.19*	4	0	.33*	5	0
D2 Knowledge of Subject Matter	.08	3	1	.11	1	0
D3 Differentiated Perceptions ¹	.04	0	0	.24*	6	0
P1 Goal Orientation ¹	-.17	0	0	-.01	1	1
P2 Program Change Based on Need	.01	1	1	.08	0	0
P3 Flexibility of Curriculum ¹	-.04	0	0	-.07	0	0
G1 Global Rating (Like Class?)	-.04	0	0	.03	0	0
A Appropriateness of Curriculum for Individuals	.02	0	0	.21*	3	0
G2 Perceived Academic Competence ¹	.10	1	0	.28*	5	0
G3 Orientation Toward Affect ¹	-.15	0	2	-.12	0	0
G4 Learning Environment (Cooperation, Student Responsibility)	.30*	3	0	.24*	5	0

* A correlation $\geq .18$ is significant at the .05 level with 120 degrees of freedom. All planning and general variables are class-level variables (25 classes in grade 2). Using student-level degrees of freedom is a liberal interpretation (see discussion in text).

¹These rating variables are residual scores with the global rating partialled out.

Table 7-4

Summary of Correlations Between Planning or
General Variables and Residual Achievement

(Grade 5)

Variable	Grade 5 Reading			Grade 5 Mathematics		
	Correlation with A-C Residual Ach. (Total Test)	Number of Significant Correlations (out of 4) *		Correlation with A-C Residual Ach. (Total Test)	Number of Significant Correlations (out of 8) *	
		Positive	Negative		Positive	Negative
D1 Item Prediction	.18*	2	0	-.16	0	0
D2 Knowledge of Subject Matter	.07	1	0	.05	0	0
D3 Differentiated Perceptions ¹	.05	0	0	-.21*	0	3
P1 Goal Orientation ¹	.22*	2	0	.02	0	0
P2 Program Change Based on Need	.08	1	1	.05	1	0
P3 Flexibility of Curriculum ¹	.07	0	0	-.09	0	1
G1 Global Rating (Like Class?)	-.08	0	0	.01	1	1
A Appropriateness of Curriculum for Individuals ¹	.17	0	0	-.02	0	1
G2 Perceived Academic Competence ¹	.10	0	0	.08	1	0
G3 Orientation Toward Affect ¹	-.14	0	0	-.06	0	2
G4 Learning Environment (Cooperation, Student Responsibility) ¹	.15	0	0	-.02	0	1

* A correlation $\geq .18$ is significant at the .05 level with 120 degrees of freedom. All planning and general variables are class-level variables (21 classes in grade 5). Using student-level degrees of freedom is a liberal interpretation (see discussion in text).

¹These rating variables are residual scores with the global rating partialled out.

Table 7-5

Summary of Correlations Between Interactive
Teaching Behaviors and Residual Achievement

(Grade 2)

Variable	Grade 2 Reading			Grade 2 Mathematics		
	Correlation with A-C Residual Ach. (Total Test)	Number of Significant Correlations (out of 7) Positive Negative		Correlation with A-C Residual Ach. (Total Test)	Number of Significant Correlations (out of 10) Positive Negative	
Explanation-Need	.05	1	0	-.14	0	4
Explanation-Planned	.13	2	0	.07	0	0
Academic Monitoring	-.05	0	0	.18*	1	0
Academic Feedback	.19*	3	0	.22*	5	0
Total Substantive ⁺	.16	2	0	.17	0	0
Structure/Direct	-.10	0	0	.10	0	0
Task Engagement Feedback	-.04	1	0	-.04	0	0

* A correlation $\geq .18$ is significant at the .05 level with 120 degrees of freedom. Both interactive teaching behaviors and achievement are measured at the student level. Sample size at grade 2 is 139 students.

⁺ Total Substantive is the sum of explanation-need, explanation-planned, academic monitoring and academic feedback.

Table 7-6

Summary of Correlations Between Interactive
Teaching Behaviors and Residual Achievement

(Grade 5)

Variable	Grade 5 Reading			Grade 5 Mathematics		
	Correlation with A-C Residual Ach. (Total Test)	Number of Significant [*] Correlations (out of 4)		Correlation with A-C Residual Ach. (Total Test)	Number of Significant [*] Correlations (out of 8)	
		Positive	Negative		Positive	Negative
Explanation-Need	.06	0	0	.07	0	0
Explanation-Planned	-.01	0	0	.01	0	0
Academic Monitoring	-.03	0	0	.03	0	0
Academic Feedback	.16	0	0	.16	1	0
Total Substantive ⁺	.10	0	0	.11	1	0
Structure/Direct	.16	0	0	-.23*	0	3
Task Engagement Feedback	-.10	0	0	-.23*	0	3

* A correlation $\geq .18$ is significant at the .05 level with 120 degrees of freedom. Both interactive teaching behaviors and achievement are measured at the student level. Sample size at grade 5 is 122 students.

⁺ Total Substantive is the sum of explanation-need, explanation-planned, academic monitoring and academic feedback.

These correlations are reported in Appendix 7-A. As a summary measure, we counted the number of "significant" correlations out of the set of correlations that were run. These two types of information are usually consistent with each other. That is, a measure with a higher correlation with total reading or total mathematics has significant correlations with more subtests than does a measure with a lower correlation with a total score. The two types of measures allow one to estimate both the magnitude and the consistency of a relationship.

We chose to interpret as indicative of real and important relationships, those correlations with residual achievement that are greater than or equal to .18. A single correlation of .18 between two student-level variables is significant at the .05 level with a sample size of 120 (or more).

Because of the large number of correlations being computed, it is likely that some correlations will be greater than or equal to .18 by chance, when there is no true relationship. That is, the experiment-wide alpha level is greater than .05. There is no sure way to know which correlations occur by chance in this data set and which are reliable and replicable. We attempted to screen on the basis of logical interpretability and on the basis of replication within this study. That is, we looked for consistent effects across grade level and subject matter. While it is possible that different factors are more or less important in different quadrants, we believe that the general principals of good instruction are likely to hold constant. We have tried to assess general functions to be fulfilled in instruction and we believe these functions should be important in both grades and both subject matter areas. Relationships that do not occur consistently, particularly those that occur in just

one quadrant, should be replicated in further studies before they are taken seriously. Relationships which occur more consistently are more likely to be real and provide a sounder basis for interpretation.

The cut-off of .18 was applied regardless of whether the teaching process variable was measured at the student level (as with the interactive teaching behaviors) or at the class level (as with planning and general variables). For class level variables, it can be argued that an adjustment should be made to reflect the smaller number of degrees of freedom at the class level. The most conservative position would be to use the class-level degrees of freedom. This would require a much stronger relationship (a correlation of about .40) to achieve statistical significance at an alpha level of .05 (for a single relationship). This means requiring a stronger relationship for class-level variables than for student-level variables in a situation where we know class-level variables are unlikely to have a truly strong relationship because of the large amount of within-class variability in achievement. We would like to be able to detect the moderate relationships that do exist. We therefore chose not to make an adjustment or use class-level degrees of freedom, but we acknowledge that this is a liberal interpretation, increasing the likelihood of some chance results. Again we look for consistency and interpretability to balance this danger.

One way that we examined the level of analysis issue was by comparing class-level and individual-level correlations for the class-level planning and general variables. In Tables 7-3 and 7-4 there are 9 positive correlations greater than or equal to .18. The average of the 9 correlations is .24. When these same correlations were rerun at the class level, the

average of the 9 correlations was .39. The correlations are larger because within-class differences in achievement have been eliminated as a source of unexplainable variance. A correlation of .39 is significant at about the .06 level with a sample size of 25 classes. A minimally acceptable correlation of .18 at the individual level can be estimated to become a correlation of about .30 at the class level. A correlation of .30 would be significant at about the .15 the level with a sample size of 25 classes. In other words, if we looked at class-level correlations for class-level variables and used an alpha level of .15 we would select about the same set of relationships to interpret as we get using an alpha level of .05 at the individual level.¹ We believe that this is a reasonable and acceptable decision rule.

Given these complexities in the statistics, it is encouraging that there are several clear trends in the data.² For the planning and general variables, it appears that item prediction, D1, and learning environment, G4, are generally positively associated with achievement. Both of these relationships appear in three quadrants but do not hold in grade 5 mathematics. This exception points out again the need for further investigation of these relationships, but there is reasonable evidence that these two factors can contribute to better achievement. There are trends, but

¹Using the .15 alpha level cutoff for class level correlations, 8 of the 9 correlations would remain significant, and two additional correlations would become significant. In all cases the shifts influence the consistency of effects across quadrants but not the identification of which variables are important.

²When interpreting correlations with achievement, it is important to consider how the teaching process measures correlate with each other. A complete set of intercorrelations of teaching process measures is given in Appendix 7-B.

less strong, toward a positive relationship between appropriateness, A, and achievement and toward a negative relationship between orientation toward affect, G3, and achievement.

All four of these variables will be discussed in detail. One aspect of the discussion will be to consider how these variables might relate to achievement, including the role of ALT facets as mediating variables. As background for this discussion, Table 7-7 reports correlations between selected teaching process variables and facets of ALT. Much of this information overlaps the data reported in chapters 5 and 6, but some is new. Unlike earlier data, ALT facets are defined here over the whole A-C period. Correlations thus tend to reflect an average of the A-B and B-C correlations, with the B-C period weighted more heavily.

Item prediction represents the domain of teacher diagnosis. The teacher is asked to estimate which items of a selected set from the BTES achievement battery the student will answer correctly. When teachers are more successful at this task, their students tend to perform better on the tests. This item prediction task is one measure of how well the teacher knows what the students can or cannot do. This task presumably requires both knowledge of subject matter, in order to analyze test items, and awareness of differences between students, in order to know who can do what. There is some evidence to support this task analysis in the correlations among the three diagnosis variables (see Chapter 6). Notice also, in Tables 7-3 and 7-4, that the other diagnosis measures, knowledge of subject matter and differentiated perceptions, tend to follow along with item prediction in relationship to achievement, but the relationships are less strong. The item prediction measure appears to be a good measure of the successful integration of several facets of teacher diagnosis.

Table 7-7

Correlations of Selected Teaching Process Variables with Facets of ALT

ALT Facet	Quadrant	Teaching Process Variable				
		Item Prediction	Appropriateness ¹	Academic Feedback	Affect ¹	Learning Environ ¹
Allocated Time	2R	14	17	12	-12	-09
	2M	-09	12	-06	-02	-19*
	5R	39*	09	43*	-71*	-05
	5M	-13	06	07	-35*	-27*
% Engaged	2R	19*	09	46*	-16	-12
	2M	05	04	24*	-26*	04
	5R	-12	06	38*	-42*	26*
	5M	28*	-17	42*	-34*	08
% Low Error Rate	2R	13	11	-05	07	-07
	2M	-13	-08	-13	-30*	23*
	5R	-10	-18*	10	24*	-24*
	5M	-09	-21*	06	15	-05
% High Error Rate	2R	-15	-15	-15	14	-08
	2M	-06	-02	00	29*	-06
	5R	-20*	-35*	03	33*	-25*
	5M	03	-20*	09	05	-03

*All correlations were carried out at the student level. A correlation $\geq .18$ is significant at the .05 level with 120 degrees of freedom. All ALT facets and academic feedback are measured at the student level. Other teaching process variables are measured at the class level, so using student-level degrees of freedom is a liberal interpretation.

¹These rating variables are residual scores with the global rating partialled out.

We would expect that teacher diagnosis influences achievement by influencing the appropriateness of tasks assigned to students. This should be reflected, at least in part, in a relationship between teacher diagnosis and student error rate. Teachers who are knowledgeable about what students can and cannot do should be better able to assign tasks (and give explanations) so that students make few errors. In Chapter 6 we found that there is some evidence of a relationship between item prediction and % high error rate. Especially in the early part of the year (the A-B data), students tended to have more time coded as "hard" in classes where the teacher was less accurate in diagnosis. This relationship was less strong in the B-C period and reversed direction for grade 5 math. Table 7-7 shows correlations for the A-C period. Since B-C values outweigh A-B values in these calculations, the relationship between item prediction and % high error rate is only moderate but still evident, except for grade 5 math. Notice that the relationship between item prediction and achievement also reverses direction in grade 5 math. The relationship between item prediction and achievement and the consistent deviancy of grade 5 math lend support to the general model of diagnosis, appropriate task assignment, successful learning activities, and achievement. It appears, however, that the student error rate variable picks up only part of this relationship. Teachers who are good diagnosticians assign fewer tasks that are extremely hard for students. Beyond this, differences in diagnostic skill are not now picked up by the error rate measure.

Notice that another variable reflecting the complex and subtle process of diagnosis and prescription also tends to be positively related to achievement (in two of four quadrants). This is the rating variable of "appropriateness," estimating the responsiveness of the instructional

program to the needs and level of readiness of individual students. While such a rating is a subjective judgment, it is based on extensive observation of instruction. It examines particularly the pacing of instruction ("Is the fast learner allowed to proceed while the slow learner works with a subject until s/he has mastered it?" Are groups flexible, responsive to change in students?). This variable reflects the degree to which diagnosis, knowledge of students, leads to appropriate prescription for each individual.

Appropriateness might also be expected to influence student achievement by influencing error rate. Adapting instruction to student needs on an ongoing basis should eliminate the occurrence of hard tasks and might increase the amount of time when students make few errors. Table 7-7 shows that appropriateness is related to student error rate, especially high error rate. In classes where tasks are judged to be more appropriate, students tend to spend less time on tasks at a high error rate. This is what we would expect. Giving students tasks they cannot do seems intuitively to be inappropriate. Notice the particularly high negative correlation between appropriateness and high error rate in grade 5 reading. This may partially explain the difference between correlations with achievement in reading and math at grade 5. In sum, both diagnosis and use of diagnosis to prescribe appropriate instruction tend to be related to achievement through the elimination of overly difficult tasks.

Appropriateness, like item prediction, has some relationship to achievement other than that mediated by high error rate. Some insight can be gained by considering the odd negative correlation between appropriateness and low error rate in grade 5. In classes where tasks are judged to be more appropriate, students spent less time on tasks where they had a low error

rate. This pattern of results was reported and discussed in Chapter 6. It may be that raters were predisposed to consider medium error rate appropriate because that was the BTES prediction at the time data were collected. It appears though that both observers and teachers are inclined to think medium error rate is more appropriate in grade 5 than in grade 2. Especially in grade 5 reading, appropriateness is highly correlated with academic focus and student responsibility for academic achievement. Some teachers in grade 5 stress academics and expect students to be able to cope with challenging tasks. Appropriateness in grade 5 reading may be associated with achievement partly because of its link with academic press and in spite of the general negative effect of medium error rate.

These relationships are complex and not fully interpretable. Planning functions are related to achievement. Part, but only part, of the relationship is mediated by error rate.

The general variable with the strongest and most consistent relationship to achievement is the weekly rating variable of learning environment, G4. This variable is a combination of two ratings, one on "cooperation" and the other on "student responsibility." Achievement is higher, especially in grade 2, in classes with higher ratings on cooperation and student responsibility.

High cooperation is described on the rating scale form as "students help each other, share materials and space; perform tasks together." High student responsibility is described on the form as "students take major responsibility for their classwork, belongings, and classroom." These two ratings tended to be highly correlated with each other (.85 for the B-C average ratings in grade 2; .78 for the B-C average ratings in grade 5 (both unresidualized)). Therefore, the two ratings were combined to form

a variable for analysis.

In order to understand better the nature of this relationship, we looked at the two component ratings and at anecdotal descriptions of key classes. Such after-the-fact rationalization may produce useful insights, but these interpretations go considerably beyond the data. Further research is needed to test these interpretations.

Cooperation and student responsibility might theoretically relate to achievement through student motivation. Cooperation and student responsibility can be valued in their own right as aspects of positive social interaction and maturity. Pursued as social values, these characteristics might or might not foster short term achievement in basic skills. For instance, when students work together they may be more inclined to socialize than when they work independently, thus lowering engagement. Or, student responsibility could describe a situation where students work largely in independent seatwork on "contracts." Again, if poorly managed, this could be a low engagement situation. On the positive side, where students work together to reach academic goals which they are expected to achieve, then cooperation and student responsibility might mean high motivation, higher engagement rates, and higher achievement.

Table 7-7 shows the correlations between learning environment and ALT facets, including engagement rate. While there is no strong consistent relationship between learning environment and engagement, there is a significant positive correlation in grade 5 reading.

Further insight into the relationship can be gained by looking at descriptions of selected classes. In grade 2, the class with the highest residual achievement in both reading and mathematics, class 17, was also the class with the highest score on learning environment. This class is

described in detail in Chapter 8. In their open-ended debriefing comments about this class, field workers noted that the teacher "stressed academics" and was "very demanding." At the same time, there was a positive tone to the classroom; the students were "enthusiastic" and "eager to please the teacher." Field workers commented that "students worked cooperatively with one another and with the teacher" and that the students "have all learned to be very responsible for their own work."

In contrast to class 17, class 1 in second grade was also high in learning environment but generally low in residual achievement. In class 1 the teacher "fosters independence by giving each child a chance at handling different classroom responsibilities such as taking roll." There is also an emphasis on positive self-concepts. Teacher and children share and compliment one another. But this same teacher "avoids" teaching math and spends much reading time on creative writing. Unlike class 17, discussed above, this class seems to value the social interactions for their own sake. Thus, in class 1, the cooperative-responsible environment is not mobilized toward reaching academic goals in the same way as class 17.

A final example from a fifth grade class gives a slightly different perspective. Class 12 in fifth grade had high residual achievement, especially in reading, and a high rating on learning environment. This teacher is also described as emphasizing academics and being demanding of students. Unlike class 17 in second grade, there is no compensating sense of positive tone or enthusiasm. The teacher is "strict" and the atmosphere is sometimes "hostile." But, the teacher "holds the students responsible for their work," checks papers and requires that they be corrected. Here one gets a stronger sense of the role of positive expectations -- the students are expected to work and learn. Analyzing the

components of learning environment separately (see Appendix 7-B), it appears that student responsibility rather than cooperation produces the correlation with achievement. That fits the description of this class -- responsibly work oriented but without human spirit. One wonders about the long-term consequences. It should be noted, however, that this class showed a more positive attitude toward reading at the end of the year than at the beginning of the year.

The second general variable which shows some relationship with achievement is orientation toward affect, G3. While the correlation never reaches statistical significance, it is consistent across quadrants. Classes with a higher orientation toward affect tend to have lower achievement. This is consistent with the finding in Chapter 5 that a stronger orientation toward affect is associated with lower engagement rates. The mediating role of ALT is further indicated in Table 7-7. Classes with a higher orientation toward affect spend less time on reading and math and have more tasks with a high error rate. As with the distinction between classes 17 and 1 above, the issue appears to be one of priorities and focus. When the teacher and students turn their attention primarily to affective concerns, this leads away from academic instruction. If teachers focus on social characteristics such as the background and personalities of their students, rather than cognitive performance, they may be unaware of or willing to tolerate high error rate.

One interactive teaching behavior tends to be consistently positively correlated with residual achievement. This variable is academic feedback. It is the only variable to have a reasonably strong positive association ($r=.16$) with achievement in grade 5 math; and the relationship is as strong or stronger in the other three quadrants.

Part of the relationship between feedback and achievement is mediated by student engagement. In Chapter 5, we reported that feedback was positively correlated with engagement in all four quadrants during the B-C period and was especially highly related in reading, where more feedback is given. In Table 7-7, showing A-C correlations, it is apparent that the strongest and most consistent relationship between a teaching process measure and an ALT variable is the positive correlation between academic feedback and engagement. Students pay attention more often when they receive more feedback about whether their answers are right or wrong.

This correlation with engagement is only a partial explanation, however, of the positive effects of feedback. Feedback tends to correlate more highly with residual achievement than does engagement. The dual nature of the effect of feedback is illustrated in Table 7-8.

Students in grade two were divided into three approximately equal-sized groups according to their engagement rate in reading. Similarly, academic feedback was redefined as a three-level variable based on equal divisions in the distribution of scores. These two variables combine to form a three-by-three table, with each student falling in one of the nine cells.

The relationship between feedback and engagement is shown by the sample size in each of the nine cells of Table 7-8. Notice that with a high level of feedback few students have low engagement rates ($n=4$) and half the students have high engagement rates ($n=24$). This pattern is reversed in the bottom row of the table. Students with a low level of feedback tend to have low engagement rates ($n=28$).

The independent effects of engagement and feedback on achievement are shown by the residual achievement scores in the nine cells. For the sample

Table 7-8

Conjoint Relationship of Academic Feedback
and Student Engagement to Residual Achievement
in Grade 2 Reading

		Engagement Rate			
		Low < 70%	Medium 70%-79%	High > 79%	
Academic Feedback Rate	High >18%	0.7 (n=4)	9.5 (n=19)	5.2 (n=24)	5.1
	Medium 13%-18%	-7.6 (n=13)	3.3 (n=17)	-0.4 (n=17)	-1.6
	Low <13%	-7.7 (n=28)	-5.4 (n=10)	2.5 (n=7)	-3.5
		-4.9	2.5	2.4	

Cell entries are average residual achievement in total reading over the A-C period, and, in parentheses, sample size.

Marginals are unweighted averages of the three cell means in that row or column.

as a whole, residual achievement scores must sum to zero. Least-squares linear regression analysis calculates a regression line so as to balance the distance of points above the line against the distance of points below the line. Simply put, for each student who gained five points more than predicted, some other student gained five points less than predicted. Across the sample, above average (estimated) gain balances out below average gain for a mean residual score of zero. In Table 7-8, residual achievement scores were averaged across students in each of the nine cells, to show the relative gain for students in each cell.

Students with low engagement rates ($< 70\%$) tended to have below average gain. Equally weighting each of the three levels of feedback, the average residual score for students with low engagement rates was -4.9 . This means that these students, on the average, gained 4.9 fewer items from the A test to the C test than the average for the sample as a whole. The overall average gain equivalent to a residual score of zero, was 48 items. Low engagement students gained about 43 items. Using the standard deviation of residual scores (21.8 items in grade 2 reading), one can express these differences in terms of percentile units. The average gain of 48 items represents the 50th percentile. Students with low engagement rates fall, on the average, at the 41st percentile in achievement gain.

In this sample in grade 2 reading, there was a major difference in achievement between students with low engagement rates and students with medium or high engagement rates. This suggests that it would be a reasonable goal to try to get students to an engagement rate of 70% or greater. From the sample sizes in each cell, it can be seen that almost all students who received high amounts of academic feedback had medium or high engagement

rates. To achieve such a high level of feedback (more than 18% of the student's instructional time feedback was presented) it is probable that some time was spent in a group setting, where students hear correct answers and thus get feedback. Other types of feedback probably also occur in order to get such a high percentage -- feedback during seatwork, self-correcting curriculum materials, oral reading, etc.

Students who receive large amounts of feedback have a double advantage. They are likely to have medium or high engagement rates. They also achieve more because of some other independent effect of feedback. Holding engagement constant, students who receive more feedback learn more. The difference in average residual achievement between low feedback students and high feedback students, independent of engagement, is the difference between being 3.5 items below average in gain versus being 5.1 items above average in gain. This is equivalent to a difference of 15 percentile points.

One might expect that this second effect of academic feedback would be picked up by the student error rate variable. Feedback should reduce the number of errors students make. When the student responds correctly, feedback should make him more sure of his answer and likely to remember it. When the student responds incorrectly, feedback should make the student aware that he was wrong and reduce perseveration of errors. Hearing the right answer gives the student an opportunity to learn the correct response.

Amount of feedback was not related to student error rate as we measured it. As shown in Table 7-7, feedback was not correlated with error rate. A plot of the grade 2 reading data was made using procedures like those in Table 7-8. This plot indicated that amount of feedback had about the same linear relationship to achievement in each of three overall levels of low error rate. (There was a slight tendency for high amounts of feedback

(> 18%) to have a greater effect for students in the middle range of low error rate.)

Several factors may contribute to the observed zero correlation between feedback and low error rate. For one thing, teachers may tend to give more feedback to students who make more errors. They may place a higher priority on correcting errors than on reinforcing correct responses. This would produce a negative correlation between feedback and low error rate. Perhaps the two effects counterbalance each other.

This explanation is a reasonable possibility. However, many of the differences in the amount of feedback children receive are differences between classes. Either the teacher has a routine of correcting papers or s/he does not. Class routines should produce a generally positive effect. Here it is important to consider the many behaviors we coded as feedback. Group lessons with a question-and-answer format provide a large amount of feedback. The student is continually hearing the correct response, from another student or from the teacher. To the extent that amount of feedback is synonymous with amount of groupwork, the relationship of feedback to error rate takes on a somewhat different meaning. One common belief about teaching is that you should introduce new material in a group setting, where examples and explanations can be given, questions asked, etc. During seatwork, where students are on their own, tasks should be easier. Students should know what they are doing by the time they get to seatwork. If some teachers followed this pattern, then more feedback would be associated with more group introduction of new material and hence more medium error rate. This factor would also contribute to an overall zero relationship.

Finally, one should realize that the categories of error rate currently coded in the BTES system are very broad. The distinctions are between being

totally lost, knowing something, and knowing completely how to do the task. It may be that finer distinctions would provide a more reactive indicator of the positive effects of feedback. As it is, error rate seems to be more strongly influenced by broader structural characteristics of the classroom, such as the pacing built into the curriculum, the organizational structure of the classroom (groupwork vs. seatwork), the value the teacher places on easy versus challenging material, implementation of initial and ongoing teacher diagnosis.

Differences between achievement subtests. The previous section discussed general effects of teaching process variables on achievement in reading and mathematics. Actually, some types of achievement tests tend to show the effect of a particular teaching process variable more than other types of achievement tests. Appendix 7-A shows the correlation of each of the teaching process measures with residual achievement on each subtest. Significant results by subtest are summarized in Table 7-9. This table gives the number of significant ($\geq .18$) positive and negative correlations for each subtest out of the 11 correlations with planning/general variables and the 7 correlations with interactive teaching measures.

There is a trend for the planning and general variables (item prediction, appropriateness, learning environment) to be positively related to achievement on scales that stress complex skills or concepts. Notice that reading comprehension at both grade levels has three significant correlations, which is more than any other subtest in reading. Also, in grade two mathematics, the highest number of significant correlations is with the concept scales in place-value/numerals, word problems, and fractions. It would appear that teacher skills in diagnosing student strengths and weaknesses, assigning appropriate tasks, and fostering academically

Table 7-9

Summary of Significant Correlations with Achievement,
by Subtest

Achievement Measure	Planning/General Variables No. Significant* (out of 11)		Interactive Teaching Behaviors No. Significant* (out of 7)	
	Positive	Negative	Positive	Negative
Grade 2 Reading				
Decoding Blends and Long Vowels	1	1	3	0
Decoding Variant Consonants	2	1	1	0
Decoding Complex Patterns	2	1	3	0
Word Structure-Meaningful Units	2	0	1	0
Syllables	0	0	0	0
Total Comprehension	3	1	0	0
Total Reading	2	0	1	0
Grade 2 Mathematics				
Addition and Subtraction, No Regrouping	2	0	1	0
Speed Test	1	1	0	0
Addition and Subtraction, With Regrouping	2	0	0	0
Computational Transfer	2	0	0	1
Place Value/Numerals	4	0	1	0
Word Problems	4	0	1	1
Money	1	0	0	1
Fractions	5	0	0	0
Total Mathematics	5	0	2	0
Grade 5 Reading				
Total Word Meaning	0	0	0	0
Total Comprehension	3	0	0	0
Syllables	1	1	0	0
Total Reading	2	0	0	0
Grade 5 Mathematics				
Multiplication	1	1	0	2
Division	0	0	0	2
Fractions	0	1	0	0
Multiplication Speed Test	1	3	0	0
Total Geometry	1	3	3	0
Computational Transfer	0	0	0	0
Word Problems	0	0	0	0
Total Mathematics	0	1	0	2

*All correlations $\geq .18$ were counted as significant for this tally. A correlation $\geq .18$ is significant at the .05 level with 120 degrees of freedom.

cooperative-responsible environments are particularly important in the development of cognitively-complex conceptual learning.

Regressions Using ALT Variables and Teaching Process Variables to Predict Achievement

As a final summary analysis, we used regression to examine the combined relationship of ALT variables and teaching process variables to achievement. For each subtest, we predicted achievement at C with 10 variables -- a pretest measure at A, four ALT variables (allocated time, engagement rate, low error rate, and high error rate) and the five most potent teaching process variables (item prediction, appropriateness, academic feedback, orientation to affect, and learning environment). With so many predictor variables in a single regression, the weights for individual variables become somewhat unstable. Our interest, however, was not so much in individual variables as in the general impact of the ALT variables as a set and the teaching process variables as a set and the relative magnitude of the impact of these two sets of variables.

One purpose of this analysis was to get a general estimate of the extent to which teaching process variables influence achievement "through" ALT. In a logical analysis, it must be the case that teaching process variables influence achievement test scores by first influencing student behavior, ongoing learning in the classroom. This is the basis of the ALT model. However, the ALT variables as operationally defined in this study are only one more or less adequate set of measures of student learning behavior. At this stage of development we would expect that some of the impact of teaching process variables is mediated by ALT as measured but that other teaching process variables relate to aspects of learning not

assessed by the ALT variables as measured. Presumably, refinement of the measurement system could increase the degree to which ALT acts quantitatively as a mediator, but it is also possible that some aspects of student learning cannot be adequately observed on an ongoing basis or easily quantified in a time metric.

In order to get an estimate of the extent to which teaching process variables go "through" ALT, as measured, we looked at the unique contribution of the ALT variables as a set and of the teaching process variables as a set. This was done by running two regressions for each dependent measure. In both regressions, the pretest score was the first predictor. (This allowed us to calculate the variance left in the posttest score after pretest was taken into account. This residual variance is the maximum of what can be accounted for by any other set of variables.) In one order, ALT variables were entered before teaching process variables. The additional variance accounted for by teaching process variables when entered last is the variance uniquely associated with the teaching process variables as a set. Conversely, in the second order, teaching process variables were entered before ALT variables, to get the unique contribution of ALT variables as a set. The results of these analyses are reported in Tables 7-10 and 7-11.

Each of these tables has the same format. The first column identifies the subtest. The second column gives the percentage of the variance in the posttest which is accounted for by the pretest. This is the same as the squared correlation between pretest and posttest. The pretest and time variable used for each subtest are identified in Tables 7-1 and 7-2. The remaining columns report the percentage of residual variance in the posttest (after removing pretest) that is associated with ALT variables

Table 7-10

Summary of Regressions Using Teaching
Process Variables and ALT Variables to Predict Achievement
(Grade 2)

Subtest	% of Total Variance Accounted for by Pre ¹	% of Residual Variance Accounted for				
		When ALT First		When TB First		Total
		By ALT	By TB	By ALT	By TB	
Reading						
Decode Blends and Long Vowels	43%	21%	11%	13%	19%	32%
Decode Variant Consonants	5%	15%	8%	14%	8%	22%
Decode Complex Patterns	30%	25%	14%	20%	19%	39%
Word Structure - Meaningful Units	29%	10%	16%	12%	15%	26%
Word Structure - Syllables	1%	20%	3%	20%	3%	23%
Total Comprehension	18%	9%	16%	8%	17%	25%
Total Reading	49%	19%	17%	18%	18%	36%
Mathematics						
Addition & Subtraction, No Regrouping	20%	6%	10%	6%	11%	17%
Speed Test	20%	25%	7%	22%	10%	32%
Addition & Subtraction, with Regrouping	3%	6%	6%	6%	6%	12%
Computational Transfer	23%	12%	11%	13%	10%	23%
Place Value/ Numerals	23%	1%	13%	2%	12%	14%
Word Problems	12%	7%	14%	7%	14%	21%
Money	30%	16%	6%	12%	9%	22%
Fractions	1%	12%	11%	9%	14%	23%
Total Math	41%	11%	20%	10%	21%	31%

¹ See Table 7-1 for identification of the pretest measure and allocated time variable used in each regression.

Table 7-11

Summary of Regressions Using Teaching
Process Variables and ALT Variables to Predict Achievement
(Grade 5)

Subtest	% of Total Variance Accounted for by Pre ¹	% of Residual Variance Accounted for				
		When ALT First		When TB First		Total
		By ALT	By TB	By ALT	By TB	
Reading						
Total Word Meaning	53%	10%	6%	9%	8%	17%
Total Comprehension	57%	10%	10%	9%	11%	20%
Syllables	34%	8%	1%	5%	4%	9%
Total Reading	70%	7%	7%	5%	9%	14%
Mathematics						
Multiplication	10%	18%	1%	17%	2%	19%
Division	19%	7%	5%	7%	5%	12%
Fractions	10%	34%	1%	27%	7%	34%
Multiplication Speed Test	53%	9%	11%	4%	16%	20%
Total Geometry	13%	10%	11%	5%	16%	21%
Computational Transfer	42%	13%	2%	11%	3%	15%
Word Problems	27%	4%	5%	3%	7%	9%
Total Math	61%	11%	8%	9%	10%	19%

¹ See Table 7-2 for identification of the pretest measure and allocated time variable used in each regression.

and teaching process variables. The unique variance associated with the set of teaching process variables is found in the column labeled "By TB," "When ALT First." The unique variance associated with ALT variables as a set is found in the next column -- "By ALT," "When TB First."

If the ALT model were totally upheld and the measurement system were perfect, that is, if all of the effect of teaching process variables on achievement was accomplished by first influencing ALT, then the variance associated uniquely with teaching behaviors would be zero. If entered first, teaching process variables would account for part of the variance in achievement but not all, since the variance jointly attributable to ALT and teaching variables would be assigned to the variables entered first. Because teaching behaviors have an indirect effect on achievement, they would not provide a maximum prediction. There would be variance left over to attribute uniquely to ALT variables. ALT variables when entered first would provide a maximum prediction. Both unique and shared variance would be attributed to ALT. There would be no unique variance associated with teaching process variables.

The other extreme is that teaching process variables and ALT variables could be unrelated to each other and could contribute independently in predicting achievement. In this case, all variance would be unique and it would make no difference which set of variables was entered first.

The results in Tables 7-10 and 7-11 are somewhere between these two positions but closer to the second. In most cases, ALT variables and teaching process variables tend to contribute relatively independently to the prediction of achievement. In addition, ALT variables account for about the same amount of variance as teaching process variables.

There are some clear examples of overlapping predictive variance, such as

Decoding Blends and Long Vowels in grade two reading. Here it makes a difference which set of variables is entered first, indicating that there is shared variance of ALT and teaching process variables which is associated with achievement. By adding the unique variances (11% plus 13%) and subtracting this sum from the total amount of residual variance accounted for by all nine variables (32%), we find that about 8% of the variance in residual achievement is accounted for by the overlap between the two sets. Part of this shared variance is the association between academic feedback and student engagement. Other parts reflect the association between item prediction and low and high error rates and the association between orientation toward affect and engagement.

In most cases, however, there is less variance associated with the overlap of teaching processes and ALT. For total reading and total mathematics, about 1% or 2% of the variance in residual achievement is associated jointly with teaching processes and ALT.

It is possible to interpret this low overlap as support for the ALT model. This argument involves recognizing a bias in the way teaching process measures were selected for the summary regressions. Consider what you would predict on the basis of the ALT model. The model predicts, and the data indicate, that some teaching variables may influence achievement solely by influencing a single facet of ALT. For instance, amount of substantive interaction is related to student engagement. In this situation, the teaching variable will necessarily have a lower correlation with residual achievement than the ALT facet has with achievement! This is true because a number of variables influence a single facet of ALT, just as many facets of ALT relate to achievement. If engagement, as measured, accounts for 4% of the variance in achievement, and substantive interaction accounts for

only one part of that 4%, the relationship will not reach statistical significance. In effect, those variables which do influence achievement by first influencing a single facet of ALT as measured will show such attenuated relationships with achievement that they will not be chosen for the summary analysis. Only those teaching variables that influence achievement in ways not now measured by our ALT variables will tend to be chosen. This is fair in the sense that these variables do relate to achievement and point out aspects of teaching and learning which are not now assessed by the student behavior variables we measure. It underestimates, however, the role of teaching behaviors that do have an impact by influencing ALT.

It is important to recognize the role of error of measurement. The true relationship between ALT facets and achievement is undoubtedly higher than the observed relationship, but the observed relationship gives an idea of what it is reasonable to expect with data of this type. The interactive teaching behaviors are handicapped by lower reliability of measurement than the reliability of engagement. They are much more infrequent than a continuous student behavior like engagement. For relationship to achievement, additional sampling error enters in (sampling of days) and makes the reliability of interactive teaching behaviors even lower. It is impressive that academic feedback does so well. The extent to which interactive teaching behaviors influence achievement through engagement is undoubtedly underestimated because of reliability.

In sum, the facets of ALT are empirically validated and immanently sensible outcome measures. ALT as measured in BTES does not include all important aspects of student learning behavior. At this stage of development, this is hardly surprising. The ALT variables do allow us to identify useful instructional variables that are not picked up by direct

correlations with achievement. They allow us to identify ways in which teachers create reasonable learning environments for their children. Both ALT and achievement are incomplete but useful measures of student learning.

Summary

Both academic learning time and performance on achievement tests are indicators of student learning. Neither one is an exact measure, since learning is an individual internal cognitive process. Each measure has advantages and disadvantages. Academic learning time has the advantage of being a relatively unobtrusive, ongoing indicator, as we have discussed before. Academic learning time as measured in this study is one of the first attempts to operationalize an ongoing measure of learning. As a relatively primitive construct, it undoubtedly suffers from measurement error and from limitations in definition. Nevertheless, we have seen that the facets of ALT are related to performance on achievement tests, the more traditional measure of learning. We have also seen that teaching process measures are related to ALT. In this chapter, we examined the relationships between teaching process measures and student performance on achievement tests. The central question was whether the same teaching process variables that were related to ALT would also be related to this second measure of learning, that is, achievement test scores. If so, we would advance our understanding of how variables relate to achievement by knowing their relationship to mediating student behaviors. If not, we would gain insight into limitations in one or both measures of learning.

As a first step, we looked at the correlations between A-C residual achievement scores and each of the measures of planning, interactive teaching

behavior and general context introduced in the last two chapters. All correlations were calculated at the student level. All correlations $\geq .18$ were considered statistically significant. For those teaching process measures that were related to achievement, we also calculated their correlations with the four ALT variables (allocated time, engagement rate, low error rate, and high error rate). A major part of the chapter is a discussion of the meaning of the relationships which occurred. Based on inferential interpretation of the correlations, we speculated on how variables influence achievement, both through the student behaviors we observed and in other ways. The most consistent and meaningful relationships with achievement were the following:

1. Amount of academic feedback was positively correlated with achievement. Students who received more feedback on the correctness of their answers learned more during the year. The correlation was significant in both reading ($r=.19$) and math ($r=.22$) in grade 2 and nearly significant ($r=.16$) in both areas in grade 5. Part of this relationship may be attributed to the fact that students who receive more academic feedback have higher rates of engagement. One might also expect that students who receive more feedback would make fewer errors. Error rate as measured in this study was not related to amount of academic feedback, and the effects of feedback on achievement were about the same regardless of the student's general error rate level. A number of possible explanations regarding error rate were offered.
2. The teacher's accuracy of diagnosis, as measured by item prediction, was positively correlated with achievement in three of the four quadrants. When teachers were more accurate in estimating how students would do on representative BTES test items, students learned more.

Grade 5 math was the exception to this pattern. The relationship between item prediction and achievement was mediated in part by high error rate. Teachers who were more accurate in diagnosis of students' academic skills were less likely to assign tasks that the students were unable to do.

3. The rated appropriateness of the instructional program for individual student needs was positively correlated with achievement in two of the four quadrants ($r=.21$ in grade 2 math; $r=.17$ in grade 5 reading). When ongoing diagnosis of students' cognitive skills was used to guide prescription throughout the year, students learned more. Part of this relationship may be explained by the fact that classes that were rated more appropriate had less high error rate. Grade 5 classes rated high on appropriateness also had less low error rate. This is counter to the current model. It appears that grade 5 teachers who stress academics may consider it appropriate to give fifth graders more challenging material. Apparently some aspect of academic press or appropriateness can influence learning independent of error rate.
4. The context variable "learning environment" was positively correlated with achievement in three of the four quadrants. The relationship was strong in grade 2 and moderate in grade 5 reading ($r=.15$). Learning environment is a combination of the rating measures "cooperation" and "student responsibility." In classes where students and teacher work together to reach academic goals that students are held responsible for achieving, learning is greater.

5. The context variable orientation toward affect was consistently negatively correlated with achievement, although the magnitude of the correlation never reached statistical significance. (Perceived academic competence, part of which is cognitive task orientation, was consistently positively correlated with achievement and reached significance in grade 2 math.)

Teachers who placed primary emphasis on the affective domain tended to spend less time on reading and math. They had classes with lower engagement rates during academic instruction. They sometimes had more low error rate tasks, but they also had more high error rate tasks, suggesting a tendency to give easy work without really considering individual student cognitive needs. When the teacher's primary attention was directed toward affect rather than academic instruction, achievement was lower.

As a summary analysis, multiple regressions were run using both ALT variables and teaching process variables to predict achievement. The five teaching process variables just discussed were selected to represent the teaching process domain. Variables were entered into the regression in two different orders -- ALT variables first and teaching process variables first. This was done so that we could estimate the unique contribution of each set of variables. The unique contribution of teaching process variables would be small if these variables had an impact on achievement by influencing ALT as measured.

These regression analyses indicated that, in general, ALT and teaching process variables contributed relatively independently to the prediction of

achievement. The shared variance of ALT and teaching process variables accounted for about 1% of residual achievement in grade 2 total reading and total math. It accounted for 2% of residual achievement in grade 5 total reading and total math. Each set uniquely accounted for 5% to 20%, with ALT and teaching process measures about equally important.

Overall, these results can be taken to mean that both ALT and achievement are useful measures of learning and that each allows us to understand some aspects of the teaching-learning process. The ALT variable of student engagement rate, for instance, reflects the positive impact of interactive teaching behaviors and provides strong evidence for this one element of teaching. Without ALT as an outcome measure, this would not be perceived. Yearlong achievement growth seems to relate especially well to the complex, high-inference rating variables like learning environment and appropriateness. Considering all three sets jointly allows us to speculate about how variables might influence learning and about how we could expand our observation of ongoing learning behaviors to reveal even more of the teaching-learning process.

CHAPTER 8: A QUANTITATIVE AND QUALITATIVE EXAMINATION OF SELECTED SECOND GRADE CLASSES

This chapter describes a selected set of second grade classes. The classes were selected on the basis of their residualized achievement scores in reading and in mathematics. The residualized scores are defined as the difference between the actual class score at the spring testing session and the score predicted for the spring testing session from the class score at the fall testing session. Classes who score higher at the spring testing than predicted from the fall testing have positive residualized scores, an indicator of greater than average change or growth. Likewise classes that score lower at the spring testing than predicted from the fall session have negative residual scores, an indicator of less than average change or growth. The residualized score can be interpreted as reflecting greater than average, average, or less than average change relative to the group of classes as a whole. Three of the selected classes showed very high residualized achievement scores on both reading and mathematics. They are contrasted with three classes that showed large negative residualized reading and mathematics scores. These two groups of classes may be referred to as positive and negative outliers. Two classes that had "average" residualized reading and mathematics mean achievement are also included in this analysis for contrast with the positive and negative outlier classes.

Correlates of the "positive", "negative" and "average" classifications will be examined, primarily in a descriptive manner. The methodology in this chapter differs from that reported in earlier chapters of the report in a number of ways. First, there is an attempt to look at multi-dimensional outcomes. Both reading and mathematics achievement are considered simultaneously. Reading and mathematics attitude data are also

examined to supplement our conception of the appropriate outcomes of instruction. Second, the analytic procedures in this chapter rely, in part, on descriptions of the classes. No statistical analyses of these descriptive dimensions will be reported.

The approach used in this chapter might be considered as a form of detective work. An event is detected, i.e., a class is identified in a relative sense to be very high in residualized means on reading and mathematics. We now pursue the following question--from all information we can examine, how is the class different (or the same) from classes that showed average or below average residualized change in reading and mathematics?

Purpose of this Form of Inquiry

In a research study such as this one specific behaviors, situations and impressions of the classroom are categorized by observers into pre-defined categories. The observers record instances of feedback, for example, following our operational definition, because feedback has been hypothesized to be an important characteristic of "good" instructional systems. Information about the specific type of feedback, whether it is oral or written, from programmed text, or obtained when the class checks papers, etc., is either lost or too complex to analyze meaningfully. The style of research used in this study is dependent on an a priori understanding of the classroom, however imperfect that understanding may be. There is an assumption that one can state ahead of time what will be important characteristics of instruction. The effects of those variables are then examined. This style of research is one that is used most often when many different observers are entering many different classes to investigate many factors of instruction. Prior training of observers, standardization and reliability are essential ingredients for judging the adequacy

of such a study, and this contributes to a continuation of this style of research.

The variables used in the BTES study, e.g., "feedback," are a bit removed from the classroom. These variables are relatively abstract, "functional" variables. Variables were purposely chosen that could describe instruction at this level of abstraction. The guiding belief was that certain functions of instruction were most likely to occur across classrooms while particular, more molecular behaviors were more likely to be classroom specific. Diagnosis, another of the functional variables used in this study, may be an essential teaching activity, but may be carried out in any number of ways. For example, it may or may not be necessary for a teacher to do formal testing when diagnosing. Research studies of classroom instruction must not always focus on the particulars, for they may then fail to detect the importance of the general function. But in focussing on the more general function, particulars often get ignored.

When this happens, the process of abstraction must be reversed if research findings about functions are to be implemented in classrooms. A teacher who wants to build more feedback into her program must find specific ways of doing so. A teacher who wants to monitor student engagement needs some guidelines as to what constitutes "high" engagement versus "low" engagement. One purpose of this chapter is to provide some specific and concrete illustrations of the more abstract constructs and functions analyzed in this study.

The format of this chapter is to describe in detail eight second grade classes, chosen to represent high, medium, and low levels of residualized achievement. Each of the class scores on the BTES variables are presented. This allows one to see the specific levels of variables and

patterns of variables in these specially selected classes. We have also tried to go into the informal records of the study and provide a richer and more meaningful description of life in these classrooms. Daily schedules, teacher verbal reports, and field worker notes are integrated to provide an approximation of case study methodology. Sometimes this approach may produce insights about variables that are characteristic of effective and less effective instruction.

Some words of caution are appropriate. The data reported in this chapter were not gathered to provide extensive anecdotal or concrete descriptive information. Thus there is a limit to the amount of detail that can be provided. In addition, in providing illustrations, the process is highly subjective requiring a good deal of inference about goals and behavioral intentions. The way a particular teacher does things may or may not be a good example of the different abstract concepts. In fact, it is a considerable act of inference to say that a particular behavior is an instance of a particular construct, especially where ratings of climate are involved. Virtually all the variables discussed in the descriptive analysis of classes should be seen as hypotheses rather than findings. Nevertheless, even with these caveats, we believe it is important to write about what the results of this study may mean in terms of the practice of teaching. We hope this will promote greater understanding of the project. These descriptions may be a starting point for teachers and teacher educators who plan to adapt and implement the BTES model of instruction in individual classrooms.

Methodology

This section describes the procedures for identifying the classes reported in these analyses.

Residualized achievement measures. For each target student residualized reading and mathematics achievement measures were computed. Using data for the 139 target students, regression equations were used to predict a total reading score at the C testing period from the total reading score at the A testing period. Similarly, a total mathematics score at the C testing period was predicted from the total mathematics score at the A testing period. Parallel procedures were used to predict reading and mathematics attitude at the C testing occasion. A residual score was thus defined for each student as the difference between the predicted score at the C testing and the observed score at that time. The mean, median, and percent of positive residuals for each class were examined. We sometimes observed a skewed distribution in a class. For example, class J had a residual mean on reading of 2.08 with a median of -6. These indices are related but each gives a slightly different definition of effectiveness. The same procedure was used in estimating residualized class effectiveness indicators for mathematics achievement and reading and mathematics attitude. Finally, classes were chosen which were the most consistent across the different indices of effectiveness.

Table 8-1 presents the residual means for reading achievement, mathematics achievement, reading attitude, mathematics attitude, and the number of second graders in our sample of 25 classes. Note again, however, that the means are based only on information obtained from about six target students per class. Furthermore, the number of second grade students per class does not always represent the total number of students in the class. Classes L and Q, for example, were split classes (a grade 1 and 2 combination and a grade 2 and 3 combination). The number of second grade students in a class appears to be a discriminating variable and will be discussed below.

Table 8-1

Number of Second Grade Students and Class Mean Residualized Achievement
and Attitude Scores (Reading and Mathematics) for 25 Classes

Class	Number of 2nd Grade Students	Mean Residualized Reading Achievement	Mean Residualized Mathematics Achievement	Mean Residualized Reading Attitude	Mean Residualized Mathematics Attitude
A	24	-12.811	-7.356	-0.0513	1.2857
B	26	-7.621	-2.534	0.1217	-1.3902
C	27	-2.322	14.740	-0.4318	0.6318
D	27	-6.194	-8.734	-2.5875	2.4423
E	25	-6.476	4.031	-0.0922	0.2240
F	29	-11.967	-38.253	-2.0167	-4.3938
G	21	5.721	10.006	-3.4468	-0.9594
H	23	-13.105	-9.081	3.1248	0.6678
I	30	-4.077	15.006	1.0248	-2.3462
J	28	2.082	-25.533	2.1752	-1.0117
K	15	22.987	-3.788	1.0933	-1.9390
L	17	20.336	8.812	-2.2148	0.4938
M	22	-13.881	-6.250	1.1154	-0.3572
N	21	-10.012	-7.070	-1.2745	-0.6728
O	34	-11.205	-13.754	-1.0418	1.7203
P	35	13.357	-3.567	3.8814	3.5710
Q	18	26.440	28.304	1.2978	1.2618
R	30	9.942	16.379	-5.8532	-5.5678
S	21	0.903	4.066	0.5242	1.8870
T	22	-4.769	-16.481	-1.3450	-1.6224
U	14	2.945	10.638	0.8607	1.4450
V	27	10.250	27.653	1.2852	0.2185
W	22	5.360	7.505	2.1940	0.4395
X	30	-18.233	3.884	0.5380	4.1864
Y	24	-3.625	-10.863	0.2098	0.4258
	$\bar{X}$	-0.24	-0.09	-0.04	0.02
	σ	12.11	15.41	2.13	2.20
	20%ile	-11.97	-10.86	-3.10	-1.62
	50%ile	-3.82	-3.05	0.16	0.31
	80%ile	9.94	10.64	1.29	1.45

Outliers. After computing residualized scores the search for outliers and classes that would serve as contrast to the outliers could begin. Three categories of classes were wanted for this analysis: (1) positive outliers on both reading and mathematics, (2) classes that were approximately "average" on both measures, and (3) negative outliers on reading and mathematics.

We found numerous classes where an outlier was detected on one variable but not another but they didn't meet our definition for inclusion. Classes R and X were at first considered interesting, but were subsequently removed from consideration because both classes had long-term substitute teachers over the course of the year.

Class Q was the most exemplary positive case. Eventually classes L, Q, and V were selected as positive outliers on both reading and mathematics, (Category I). Classes E and S were chosen to represent average performance (Category II). Classes F, H, and O were selected as negative outliers (Category III).

The reader may wonder about the correlations between the residualized achievement and attitude measures. Table 8-2 shows the association between these measures. The residualized achievement means across the 25 classes correlate .51. The correlation between the residualized attitude measures was .48. The relationship between residual reading achievement and residual reading attitude was quite low. The relationship between residualized mathematics achievement and residualized mathematics attitude was also low. This means that different instructional variables are involved in promoting attitude from those effective in increasing achievement. Attitude measures will be examined in more detail for the classes chosen for this study.

Table 8-2

Table of Intercorrelations of Class Mean Residualized Achievement
and Attitude Scores in Reading and Mathematics
for 25 Second Grade Classes

Variable		Variable 2	Variable 3	Variable 4
1	Mean Residualized Reading Achievement	.51	.04	-.08
2	Mean Residualized Mathematics Achievement		-.01	.15
3	Mean Residualized Reading Attitude			.48
4	Mean Residualized Mathematics Attitude			

The next section of this chapter presents some quantitative data on the eight classes selected for more intensive examination.

Quantitative Data on Selected Classes

This section provides quantitative information about the selected classes. Information is given first about the change patterns in reading and mathematics for each class from the A to C testing sessions.

Tables 8-3 and 8-4 display class mean achievement data (unresidualized) and raw percent change scores for each of the eight classes. Table 8-3 displays the mean percent of items passed at the A, B, and C testing periods. The mean for a class is the average percent correct (items) for the target students in the class. It must be pointed out that the composition and number of items within tests changes over the testing sessions, especially in mathematics. Some harder items were placed in the B and C test batteries.

The class means for the A testing period may be examined to see if there are differences in the pre-measures for the three different types of classes. The classes are not comparable on the scores at the A testing. For both reading and mathematics the Category II classes are higher, on the average, than the positive and negative outlier classes. While outliers are defined by "residual" achievement it is important to recognize that almost all classes show some degree of positive growth. The percent of items passed increased from the A to C testing period in all but one case. Children are learning in almost all classes. This important point needs to be emphasized because growth resulting from schooling may be masked when the variables appear in residualized form, especially for Category III classes who are defined as negative outliers.

Table 8-4 shows an interesting pattern of comparative change from the A-B period to the B-C period. For both reading and mathematics during

Table 8-3

Class Means on Total Reading and Mathematics for the "A," "B," and "C"
Periods for 8 Selected Second Grade Classes

Category	Class Code	READING			MATHEMATICS		
		Mean % Items Passed			Mean % Items Passed		
		A	B	C	A	B	C
I	L	40	59	79	44	52	65
I	Q	28	41	75	26	40	63
I	V	28	34	64	28	33	65
II	E	55	51	72	33	36	55
II	S	48	65	72	43	46	62
III	F	17	24	41	31	24	29
	H	33	32	51	31	29	45
	O	24	28	46	39	33	49
BTES MEAN		35	42	62	37	39	55

Category I Positive Outlier, Reading and Mathematics
 II Average Outcomes, Reading and Mathematics
 III Negative Outlier, Reading and Mathematics

Table 8-4

Raw Change Scores (Reading and Mathematics) over the "A" to "C"
Period for 8 Selected Second Grade Classes

Class Code	READING			MATHEMATICS		
	Mean % Change			Mean % Change		
	A-B	B-C	Total	A-B	B-C	Total
L	19	20	39	8	13	21
Q	13	34	47	14	23	37
V	6	30	36	5	32	37
E	-4	21	17	3	19	22
S	17	7	24	3	16	19
F	7	17	24	-7	5	-2
H	-1	19	18	-2	16	14
O	4	18	22	-6	16	10
BTES MEAN	7	20	27	2	16	18

the A-B period, the Category III classes show small and sometimes negative change as compared to the Category I classes. It is possible that Category III classes take longer to get organized for instruction.¹ The Category III classes lag behind the other classes from B to C but the disparity is less.

Table 8-5 shows a large set of general classroom variables for the eight selected classes.

Classes L, Q, and S were split classes (a grade 1 and grade 2 or a grade 2 and 3 combination). The number of second grade students in these classes was smaller than found in most of the other BTES second grade classes. This issue deserves further comment. It is possible that a class of, say, 14 first graders and 14 second graders creates a different managerial task for a teacher than a class of 28 second graders. It may be that most of the first or third graders work on entirely different activities (text, workbooks, etc.) than the second graders, thereby allowing the teachers to manage two distinct learning situations. This guess is highly speculative.

The Category III classes were somewhat larger than the Category I and II classes.

For both reading and mathematics, the Category I classes had more kinds of assistance available to the teacher than typically found in the other classes. The assistance was in the form of teacher aides, adult volunteers, and/or student tutors.

¹Recent research by Carolyn Evertson (personal communication) at the Texas R & D Center and by Beatrice Ward and William Tikunoff (A Naturalistic Study of Initiation of Students Into Three Classroom Social Systems, presented at the AERA meeting, Toronto, Canada, March 1978) at the Far West Laboratory suggest that teachers differ dramatically in their handling of instruction in the first weeks of school in the fall.

Table 8-5

General Classroom Variables for 8 Selected 2nd Grade Classes

Class Code ²	Number of 2nd Grade Students	Split 1.2	Split 2.3	Number of categories of assistance available to teacher (aides, adult volunteers, and/or student tutors) in reading. ¹	Number of categories of assistance available to teacher (aides, adult volunteers, and/or student tutors) in mathematics.	Teacher Affect Score
L	17	X		3	3	47
Q	18		X	2	1	42
V	27			3	3	54
E	25			1	2	44
S	21		X	3	3	66
F	29			1	1	57
H	23			1	1	56
O	34			3	3	35
BTES MEAN (SD) (25 classes)						50 (10)

¹ This information was reported by the teacher in September and may or may not have remained constant during the year.

² Classes L, Q, and V are in Category I, positive outliers.
Classes E and S are in Category II, average achievement.
Classes F, H, and O are in Category III, negative outliers.

8-13

Table 8-5 (Continued)

General Classroom Variables for 8 Selected 2nd Grade Classes

Class Code 2	Class Learning Environment Score	Mean Wait Time 1	Mean Transition Time 1	Mean Management Time 1	Median days of school missed (B-C period) for target students
L	61	4 (1.1)	29 (1.6)	5 (1.3)	0.5
Q	70	5 (1.5)	36 (3.8)	8 (2.1)	6.0
V	53	3 (1.6)	43 (2.8)	7 (0.9)	5.0
E	33	6 (2.2)	37 (6.0)	3 (1.6)	4.5
S	58	4 (1.0)	35 (1.8)	9 (0.8)	1.0
F	54	3 (0.9)	26 (4.1)	7 (0.8)	4.0
H	53	11 (1.5)	36 (3.3)	3 (0.7)	7.0
O	53	12 (2.2)	44 (1.5)	1 (0.7)	1.5
BTES MEAN	50 (10)	4 (3.1)	34 (6.3)	6 (3.0)	N/A

¹These variables come from direct observation. Data are for the B-C period.

²Classes L, Q, and V are in Category I, positive outliers.
 Classes E and S are in Category II, average achievement.
 Classes F, H, and O are in Category III, negative outliers.

In earlier chapters the relationships of teacher affect and class learning environment to achievement was discussed. It maybe noted here that the Category I classes had, on the average, a lower rated teacher affect score than the Category II and Category III classes. But the Category I classes had a higher rated learning environment than the Category II and Category III classes. The median rated learning environment of classes in Category II was lower than for the Category III classes. Mean wait time (average time that pupils wait for an assignment to be given) was, on the average, much higher for the Category III classes. No obvious between-category differences were observed for transition time. Management time was less for the Category III classes. This was not expected. The median days of absence was slightly less for the Category II Classes.

It would appear that achievement is generally associated with a smaller number of students at a particular grade level in a class, a larger number of adult persons available to assist teachers, and a smaller amount of time that pupils must wait for the class to start work. It also appears, as in earlier analyses, that the combination of positively rated learning environments and low rated teacher affect is related to higher student achievement.

Table 8-6 presents ALT and instructional process, for reading, for the eight selected classes. Category III classes, on the average, received less daily instruction in reading than the Category I and Category II classes. The variable of percentage of time that pupils were engaged was different in the three categories. As presented earlier in the report, pupil engagement is a central variable in the ALT model of instruction. An examination of the standard deviations (in parentheses following the mean) provides further insight. The variability is smaller for the Category I than for the Category II and Category III classes. Perhaps this measure reflects the ability of

Table 8-6

Time and Instructional Process Variables (Reading) for 8 Selected 2nd Grade Classes

Class Code	Daily ¹ Time	Engagement, Percentage ¹	Low Error Rate Per- centage ¹	% Seatwork ¹	% Group Substantive Interaction ¹	% Academic Feedback ¹	Appropriateness of Instruction Score	Diagnosis (Item Prediction)
L	83 (5)	79 (3)	59 (15)	66	23	21	54	57
Q	118 (3)	74 (2)	64 (17)	69	22	21	57	59
V	86 (6)	82 (5)	32 (13)	58	29	18	57	46
E	96 (5)	69 (6)	54 (10)	51	28	18	46	59
S	90 (2)	70 (7)	58 (11)	65	15	11	57	50
F	65 (5)	71 (5)	64 (12)	51	26	21	40	31
H	75 (6)	61 (10)	47 (21)	68	11	8	30	49
O	96 (3)	57 (8)	39 (9)	71	16	11	46	24
BTES MEANS	88	73	53	63	21	16	50	50

8-16

¹These variables come from direct observation. Data are for the B-C period.

a teacher to get almost all of their students engaged at a consistent rate.

Category III classes generally received a smaller level of percent easy tasks than the Category I and Category II classes. However, class F in Category III had the same percent easy (64) as class Q, a Category I class. The Category II classes received a smaller percent of time in seat work than the Category I and Category III classes. The percent of group substantive interaction received by Category II and Category III classes is generally below the level received by Category I classes. Category I classes also generally receive more academic feedback than the other classes. The field workers, typically, gave higher ratings of appropriateness of instruction for target students to Category I and Category II teachers than to Group III teachers. Category I and Category II teachers were better able to make accurate diagnoses (item prediction scores) than Category III teachers.

Table 8.7 presents ALT and instructional process measures for mathematics instruction for the eight selected classes.

Category III classes received, on the average (median time) six minutes a day more mathematics than the Category I classes and seventeen minutes more a day than the Category II classes. However, the Category III classes were less engaged, had a lower percent of easy tasks, had a higher percent of seat work and a lower percent of group substantive interaction and academic feedback. The Category I and Category II classes had higher average appropriateness of instruction scores than the Category III classes. The diagnosis scores were all lower for the Category III classes than for the Category I and II Classes. The findings reported for mathematics generally support those reported for reading.

To supplement the quantitative descriptions prose descriptions of the three Category I and three Category III classes are now presented.

Table 8-7

Time and Instructional Process Variables (Mathematics) for 8 Selected 2nd Grade Classrooms

Class Code	Daily Time ¹	Engagement Percentage ¹	Low Error Rate Percentage ¹	% Seatwork ¹	% Group-Substantive Interaction ¹	% Academic Feedback ¹	Appropriateness of Instruction Score	Diagnosis (Item Prediction)
L	37 (4)	79 (4)	66 (14)	87	8	7	58	59
Q	32 (2)	85 (3)	73 (7)	61	29	8	44	46
V	43 (3)	79 (6)	55 (10)	52	24	10	64	55
E	29 (4)	71 (8)	67 (16)	56	28	12	44	54
S	22 (3)	62 (8)	56 (11)	68	6	4	64	70
F	24 (4)	75 (8)	51 (14)	71	13	5	29	25
H	42 (4)	63 (14)	59 (22)	87	7	2	40	29
O	43 (3)	54 (9)	37 (15)	79	11	8	44	44
BTES MEANS	36	71	51	73	14	7	50	50

¹These variables come from direct observation. Data are for the B-C period.

Descriptions of a Set of Outlier Classes

The prose that follows is a synopsis of information available from the source documents of teacher interviews, field worker reports, and perceptions of the Far West Laboratory staff. The information is designed to supplement the quantitative information discussed earlier by providing a richer description of the classes and teachers.

Class L (Category I). This class was a split one-two class. The teacher had previously taught grade 1. At least three of the target students has been in her class the year before. The students who were chosen to be in this split class were perceived by the school to be the "high" group. They were perceived to be students who could work independently. The students did have relatively high, but not exceptionally high, scores at the time of the A testing. It is interesting to note that class K was in the same school, was a split 2-3 class, and was perceived to be a "lower" group. Class K did start lower in reading (34% correct) but not in math. Class K also had very good gains in reading (in fact, higher than class L, for a total change of 43%). Thus, the perceived status of class L students in the school cannot completely explain their high achievement. In fact, in the methodology of a research study, a high entering score may be a disadvantage in that ceiling effects in the test or chance testing effects may make it less likely for initially high scorers to show gain. Thus class L did well in spite of and not because of its relatively high initial standing.

Class L was described as a relatively traditional class. The curriculum was described as "average second grade." A split reading scehdule was used, with half the class present for an hour in the morning and half present for an hour in the afternoon. Students were assigned to reading groups and

worked in a basal reader series. Students generally worked on math from 10:10 to 10:50 and on language arts from 11:15 to 11:45. The afternoon was generally spent on non-academic activities.

Especially in math, a relatively high percentage of the time was spent in seat work. A field worker reported that a typical instructional pattern would be for the teacher to assign workbook or ditto pages "'plus two in your speller' or other supplemental workbook." When the children finished their assigned activities, they were allowed free time. An important factor in making this approach work appeared to be the number of people available to monitor and provide help. The class was in the ECE program and had a paid aide, parent volunteers and high school tutors working in the classroom. In addition, students helped each other. In particular, one target student often tutored first graders and another of the target students. Field workers commented that this peer tutoring appeared effective.

The classroom atmosphere was described as relaxed. The teacher was described as pleasant and good-humored but kids who did not conform were 'called down' by the teacher. The teacher rewarded academic achievement by giving "Bionic jumps" and "Bionic handshakes," which the students were eager to receive.

Class Q (Category I). The teacher in this class was male. His previous teaching experience had been in the upper elementary grades. This year, for the first time, he had a second grade (split 2-3) class.

The class was described as "traditional". Students sat in desks in five straight rows facing the blackboard. The room was "spartan". There were few bulletin boards, although some student art work was on display. The focus in the classroom was on academic instruction. The academic program

will be discussed first.

Field workers reported that he was very organized. His planning was "thorough" and done in advance. During daily lessons he worked very closely from his plan book. Indicative of this planning, the daily schedule he reported in September was followed relatively consistently throughout the year. A day in this classroom tended to look as follows:

8:45 - 9:45 Morning Reading

Half the class was present for the morning reading session. Four of the six target students were present. Students were assigned to reading groups and each group had a basal reader textbook. In January, one of the four target students was in "Group A," which was still working in a primer ("Around the Corner" published by Harper & Row). The other three target students were in "Group B," working, in January in a grade one reader ("Real and Make Believe" published by Harper & Row). Group B met first as a group, often doing 10 minutes or so of oral reading. This group meeting was also used for comprehension questions and for discussion of seatwork assignments. The last half-hour of the period was spent in independent seatwork. Group A, the lower group, spent relatively more time working as a group with the teacher. Much of this group time was spent in oral exercises in different sub-content areas of reading.

9:45 - 9:55 Recess

10:00 - 10:30 Math

Usually about 5 minutes was spent on class business, then the grade 2 group, including all 6 target students, began doing math. Each day's lesson began with the teacher explaining the concepts, directions, and goals. The teacher "used good examples to clarify any questions." Students would then work independently on the assignment while the teacher monitored their work. The teacher "rarely sat at his desk but did a lot of monitoring to be sure the task was being done correctly and to answer any questions." At the end of the lesson, papers were corrected. Sometimes the whole group went over the assignment with each student correcting his own paper. Sometimes the teacher and student tutors provided individual feedback. The fact that all second grade students

did the same math assignments is reflected in the slightly below average rating on appropriateness and perhaps also in the slightly below average diagnosis (item prediction) score. A field worker commented that the teacher "introduced math materials very gradually until all had mastered it." This approach shows up in the high percentage of time at a low error rate. Faster students who finished daily assignments early "found something to read" until others finished.

10:30 - 10:55 Physical Education (or Handwriting or sometimes more math)

10:55 - 11:05 Recess

11:10 - 11:50 Language Arts (spelling, grammar, some creative writing)

Again whole class activities. The lesson began in a group setting with teacher explanation and direction. Students then carried out the assignment individually. The teacher and student tutors monitored and provided feedback. Sometimes several tasks would be covered during this period, in which case the groupwork-seatwork cycle would occur several times. The teacher would also call the group to attention briefly during seatwork, to ask a question or give directions, if the need arose.

11:55 - 12:55 Lunch

1:00 - 1:30 Reading Lab

All grade 2 students spent one half-hour each day in the reading program run by the Miller-Unruh reading specialist. This program has two content strands -- one a supplementary phonics skills program, the other reading in a series of books called "Appreciating your Country" (Century Communications). Each strand is highly structured and sequentially organized. Students go through the same sequence but at different rates. In the reading lab, one high school tutor works with two students. Often one of the two students must wait while the tutor works with the other student; involvement is intense when the student is actually working. The tutor keeps careful records of the number of correct and incorrect responses. The target range is 90% - 94% correct. If a student falls out of range on three successive days, his assignment is altered. Students are continuously rewarded for correct responses using chips as tokens. Review activities are built into the sequence. The Miller-Unruh reading specialist oversees this operation. He also worked directly with the "Group A" target students who began the year reading at the primer level. In addition, the specialist knew personally of the progress of other students. Field worker reactions to this program were very positive.

The slightly-below-average rated affect score in this class was due to the clear emphasis on cognitive over affective goals. The "conviviality" of the classroom itself was rated slightly above average. The purpose of school is learning but learning takes place in a supportive atmosphere. This may explain the positive (although not strongly positive) residual attitude scores in this class. They liked school better at the end of this year.

One important aspect of this classroom environment was "cooperation/student responsibility". Field workers commented that "he set a positive tone and climate which fostered the cooperative spirit the kids have with one another" and that students "have all learned to be very responsible for their own work." Although not many specific details about cooperation and student responsibility are available, some comments and inferences are possible. Students checked their work with each other. The teacher himself commented that one student looked on another student's paper "only to reaffirm his answers because he lacks confidence in his own ability." This comment indicates that the teacher knew what was going on, that he had some concern that students do their own work, but that he did not rigidly enforce independence where the purpose was feedback and reinforcement. Notice also that the structure of the program builds in accountability and teaches student responsibility within that framework. Assignments are checked at the end of each lesson. Students generally have 15 to 30 minutes to do the seatwork assignment and they know it will be checked. Further, they know the teacher really cares whether they can do it or not and that help is available.

Class V (Category I). This instructional program was somewhat more complex than classes L and Q. At the beginning of the year this teacher and another teacher teamed up to regroup students during the day so that they could create homogenous groups of students who needed instruction in specific skills. The amount of planning and transition time that were involved became intolerable and they switched back to self-contained classes. This only slightly reduced the level of complexity, however. Within the class there were usually several groups working simultaneously on different things. The class was in ECE, and aides and parent volunteers worked with small groups of students. Students were free to move about and worked at tables, on the floor, or in the corners. The teachers herself sometimes referred to the classroom as a "zoo." The field workers reported somewhat chaotic conditions. The teacher would sometimes become exasperated and make sarcastic comments to students. One field worker felt that there were too many children for the available physical space.

Several factors appear to have held this situation together: 1) the teacher had a high level of concern for the students, was flexible, had a sense of humor and a high energy level, 2) there was a reasonably high degree of underlying organization and planning, especially related to academic diagnosis and prescription. There was frequent testing/quizzing. Lessons were planned based on specific goals. The teacher knew a lot about each child: "If asked about a specific child she would go into depth about their home background and about their academic strengths and weaknesses." In fact, part of the teacher's frustration seemed to be due to the fact that she was very aware of individual student progress (or lack of progress) at a detailed level. What frustrated her was inability to get a student to grasp a concept.

This teacher was particularly successful in teaching math. Four aspects of the math program are worth noting: 1) She gave daily timed tests in addition and subtraction facts, which also provided her with diagnostic information. She included drill in her program to supplement a math curriculum which emphasized concepts. 2) During transitions such as before recess and before lunch each student had to answer correctly a basic fact problem he had been having trouble on. This oral quizzing was used as a transition routine. 3) When students reached the bull's eye in a target recording basic fact skills, they were rewarded with an ice-cream cone. 4) More math time than reading time was at a low error rate.

This teacher also, more than L and Q, was reported to include conceptually rich activities. She would hold class discussions. She had a "personality of the week" writing activity where each child had a chance to be the "personality" and all students wrote something positive about the personality. She had "many other interesting creative writing ideas" as well. She would often have relaxed times when everyone sat in a big circle on the floor and she read a story and they discussed it or they had "sharing."

Class F (Category III). This teacher was a man who had previously taught kindergarten. He was a warm, easy-going, soft-spoken person. The classroom climate was "peaceful" ("even when he was angry"). In addition to the teacher, there was an aide who was "a warm, delightful teacher loved by the children." As the year went on, she was given increasing responsibility in the classroom.

The emphasis in this classroom was on things other than achievement in basic skills. The field workers reported that the teacher placed heavy emphasis on teaching social values, was concerned with improving students' self-concept, and considered time spent on non-academic activities

very important. One-half hour was spent on music almost every day. "On days when the classroom mood became tense, he spent a large part of the day doing artwork or singing." This orientation may be a carry over of this teacher's experience and feeling from teaching kindergarten. There was very little academic press in the class. It is worth noting that this class had high negative residual scores in attitude as well as achievement. Student attitudes would presumably be important to this teacher, and attitude toward reading and math as we measured it was not fostered in this classroom.

Two instructional groups were created for mathematics. The pupils were randomly assigned to the groups. Five to six groups were created for reading and the assignment was based on the previous year's reading scores. In math, the teacher would work with one group and the aide would work with the other. They would alternate groups on different days. Especially when the aide worked with a group, much time was spent on games and developmental activities.

While the teacher was clearly concerned about individual students, his attention was not directed toward academic diagnosis or specific academic goals. He felt that "most of his students had strong personal problems." (Field worker records documented some legitimate concerns.) The teacher expressed both positive and negative expectations for the different pupils in the class, with "performing at grade level" a typical statement of the year long goal. There was no reported sense of progress build into the curriculum or experienced by the students.

Class H (Category III). This teacher had many years of teaching experience. She appeared to know what she was doing and to run a generally

balanced program. Her ways of reacting in the classroom were not always consistent -- "the teacher yelled at the students one moment and in another moment called them 'sweetheart.'" She sometimes told the students not to bother her. The students seemed to have adjusted to her style.

This was one of the few grade two classes which was not on a staggered reading schedule (see also class 0). Thus all students in the class were doing reading at the same time. The teacher had reading groups and organized lessons for the different groups. She had an aide who sometimes met with a group but was not often actively engaged in instruction. In fact, the amount of substantive academic interaction was low in this class. Notice especially the low percentage of feedback. Students apparently did not do group oral reading but sometimes read for the teacher individually. During about two-thirds of the time in the group setting in reading there was no immediate interactive instructional behavior. As discussed in chapter 5, such a pattern is associated with low engagement. Notice that this class also has a high percentage of overall wait time (11 minutes per day), as the class waits for an assignment to be given. Both factors may indicate a slow pacing of instruction.

It also appeared that the teacher was willing to tolerate lack of engagement. There was one student who consistently day-dreamed (sometimes for a full forty minute period) but the teacher didn't seem to see it. The teacher didn't notice the student not paying attention even while the student was participating in a small reading group.

In addition, as suggested by the lack of monitoring and feedback and the low appropriateness score, there was no systematic, goal-oriented structure to the curriculum. The teacher appeared to know curriculum and she established groups in reading and mathematics. However, one field worker felt that the teacher was proud of the class but was not in "tune" with what the children

were actually doing. If a student failed a task the student was told "you did it ok yesterday" but not attempt was made to reteach the skill to the student. (Instead, he was given something else to do.) There were no observed grouping changes in either reading or mathematics over the year. There is no record of any positive reward system in the class.

Class 0 (Category III). This was a team-teaching situation. Two women with many years of teaching experience had developed a cooperative teaching arrangement. During reading and math periods they regrouped students so that the "slower" half of the combined group of children went to one room and the "faster" half to the other room.² The school was a "pod" type structure, which facilitated movement between "rooms." The teachers had been using this system for several years. Of the target students, all six were in the "fast" reading group. Two of the six were in the "slow" math group, and therefore had the same teacher for reading and math. The other four were in the "fast" math group, so they had one teacher for reading and the other for math.

The instructional program of both teachers was very similar. About half the time the students worked independently in an "individualized" program. Most students went through the same structural, sequential curriculum, but they worked at their own pace. Both teachers used some group-work was a whole class lesson. In reading, children met in small groups. These groups were not directly related to the independent seat work program, but provided a second strand of instruction.

²This school did not have a staggered reading schedule, so about 30 students were in each room during all academic instruction.

This organizational approach produced a number of inefficiencies. A particular problem was the lack of connection between groupwork and seat work. Students did not get an immediate opportunity to practice the material covered in groups. Students did not get group directions about what they were to do in seat work. During seat work, the teacher was continuously involved in checking students' work and giving help; but the teacher would often have to explain the same page six times to six different students in the space of five minutes. In trying to keep all students going, the teacher had to keep contact with any one student so brief that to one field worker contacts appeared "quick and superficial." And even with continuous teacher availability, there was a high percentage of wait time (36% during reading in the A-B period).

Feedback during seat work took on a particular pattern in this situation. If a student made an error on a worksheet, the teacher would mark the problem wrong. The student was then expected to correct the error. Some students learned that by successive elimination through several rounds of guessing-feedback-guessing-feedback they would eventually get the right answer. It is doubtful with such an approach that the student either understood what was wrong or really attended to the right answer when s/he had it.

Another noteworthy characteristic of this classroom is the low percentage of low error rate in both reading and math. It appeared that the teachers felt it was appropriate for these students to be working on relatively difficult material. The fast math class was already working on regrouping during the A-B period. They spent more time on regrouping during the A-B period than any other class in our sample. This in spite of the fact that the A test scores of their target students were about average. In a fall interview, the teacher noted that one student was having trouble with regrouping and also that the student disliked math and tried to avoid it. In

spite of this perception, it apparently never occurred to the teacher to move the student to an easier assignment. It is interesting to contrast this situation with the behavior of teacher Q. In November, he moved a student from the grade 2 basal reader to the grade 1 basal reader because the student was having difficulty and he felt it would be better for the student to develop confidence and speed on the easier material. Perhaps the message is this: Emphasis on academics does not mean giving students lots of difficult academic work. Instead, it means working for the steady progress of each student.

A final comment should be made about classroom atmosphere. Both teachers stressed academic content. They were described as being from the "old school" and maintained strict discipline in following classroom routines. (It was reported that all the students in both classes could be organized for psycho-motor exercises in one minute's time.) But the teachers tended to appear cold and impersonal. One field worker commented that a certain "class spirit" was missing. During groupwork, the children sometimes appeared like robots. Only rarely was there praise for work well done. These teachers were very clearly conscientious people. They wanted to do a good job of teaching the children. They purposely tried out new teaching techniques -- team teaching, "individualized" programs, seating at tables to promote social development. One teacher was heavily involved during our observation year in writing a proposal for the school's ECE program. They worked hard at their teaching. There may have been something missing in terms of feeling for students or for the individual teacher-student relationship which these second-grade children needed.

Discussion

There are many reasons to be cautious in reviewing the data and interpretations in this chapter. The selection of the classes is somewhat arbitrary. While statistical definitions of "outlier" classes were used another investigating team might utilize different definitions and therefore select other classes from the complete set of 25 second grade classes. In addition to using quantitative measures of classroom processes from the BTES files such as time measures, test score data, field worker ratings of teaching, etc., we have also utilized a qualitative approach to better understand the classes described here. We have examined the source documents of teacher interviews and field worker evaluations and debriefing. The information in the source documents was sometimes sketchy. Many types of bias and error are expected in these documents. Different field workers have different frames of reference and points of view as to what they see, value, and report. Some teachers and field workers gave us more detailed information than others. Typically, each teacher was personally observed and interviewed by two field workers and our data for teachers reflect the specific field worker teacher combinations. Finally, we now wish to answer questions that we did not anticipate earlier, and, therefore, information is not in the files. An example of the latter might be illustrated by the question of how many tutors or aides were present in a classroom rather than knowing whether any aides or tutors were present. These points are not made to destroy the value of the documents. Important information is in the field worker reports. We are attempting to point out the limitations of some of the information available to us for our investigation. The examination of information in these documents for the classes reported here was not done in a "blind" form. The classes selected were known to be

special cases and undoubtedly bias and selective interpretation of information from these documents is highly probably. In fact discussions were held formally and informally with staff members about the classes selected. Questions were referred to staff members for information they might have about the classes, teachers, and/or pupils that might help interpret the classes' achievement and attitude performance. The approach can be criticized for its lack of total objectivity.

Despite these caveats, the investigation of positive and negative achievement outliers has been an interesting and productive activity. One wishes that time would have allowed for further investigation of other interesting outliers in the second grade set of data and that a parallel investigation could have been undertaken with the fifth grade classes.

It is interesting to note that "curriculum" (textbook series) has not represented a crucial variable in our model of classroom learning. A decade ago the comparison of outcomes as a function of curricula materials would probably have been pursued. We chose not to explore the question on the following grounds: (1) Many classes used the same textbook and without detailed knowledge of how the teacher taught the curriculum package the question cannot be answered; (2) Curriculum is of less direct importance for teacher training.

The relationship of achievement to attitude toward school subjects remains an interesting question. A few classes produced high residualized achievement and attitude means (notably class Q). Likewise a few classes produced low residualized achievement and attitude means (notably class F). Throughout the study we have acknowledged the special problem of assessing attitude at grade 2. The correlation of .48 between the two residualized attitude scores suggests we were measuring some reasonably consistent

component. The nature of student attitudes and the relationship between instruction and attitudes deserve thorough study.

It was encouraging to find that the descriptions of climates in the positive outlier classes were positive and that attitude scores generally improved slightly as well. Classrooms do not have to be punitive in order for academic learning to take place. A major conclusion from the totality of BTES second grade data is that teacher attention to academic learning must take place. Teachers who have an academic orientation, have academic goals for their pupils, assess pupil progress toward these goals, etc., frequently have classes that exceed the expected achievement levels at spring testing. These classes have higher levels of academic learning opportunities for both reading and mathematics. Pupils in these classes receive higher levels of group substantive interaction and academic feedback. From a policy perspective there appear to be many positive indicators of teaching behaviors and other classroom variables that are associated with academic growth. While these findings are speculative, and deserve further research, they serve as a partial list of variables (perhaps teacher skills is a better phrase) that trainers of teachers, both pre- and in-service, and teachers themselves should examine in terms of their potential for improving instruction.

We are convinced that there is no single attribute that always distinguishes effective and less effective teaching situations. For some of the indicators reported here, there was some uniformity in variables that distinguished positive and negative outliers. It was also observed that on some dimensions one negative outlier class more closely resembled a positive outlier than the two other negative outliers.

The phenomena of split grade classes being associated with positive residualized achievement for the classes reported here is interesting. While one can speculate on how split classes might be associated with positive achievement gains, further investigation is in order. New split classes might be studied to see if these classes are also productive and if so, how! The concept of replication and further investigation applies equally well to all of the findings reported here, however. Further insights would undoubtedly be obtained from a similar investigation of the BTES fifth grade classes.

The association of categories of assistance and residualized achievement gain also deserves further investigation. Some of these classes were in the Early Childhood Education program (ECE) and through this program had aides and parent volunteers in the classroom. Given extra time one could examine the BTES files to identify ECE classes to see whether these classes had positive residualized scores and, if not, how they were different from other ECE classes. (Use of aides may be critical.)

The variable of pupil engagement rate is a crucial dimension of the definition of academic learning time. In addition to noting that positive outlier classes had higher average engagement rates, we found these classes to be less variable around the average. Since it is unlikely that all students naturally pay attention, this suggests a teacher effect. How does the teacher affect engagement rate for the individual pupils? How much of the engagement rate is manipulable and how much might be considered as a trait of the learner, i.e., is reasonably stable across different learning situations, with different teachers, etc? Perhaps we are learning that teachers will need to attend to characteristics of the learner such as engagement as well as pupil aptitude, attitude, self concept, motivation,

and background.

Summary

Eight second grade classes were selected for an intensive, descriptive study. Three of the classes were positive outliers on achievement measures (significantly greater reading and mathematics achievement at the spring testing session than predicted from the fall testing session). Three were negative outliers (performing significantly below levels predicted from fall testing on both reading and mathematics). Two classes performed as predicted at the spring testing session. Almost all classes in the BTES study showed positive achievement growth from the fall to spring (A to C) testing sessions. The selected classes represent above average, average, and below average achievement gain within our sample. Although classes were selected on the basis of achievement scores, we also examined attitude scores for all classes.

For these eight selected classes, detailed quantitative and descriptive information was reported. The purpose of this intensive study was to supplement the statistical analysis of abstract variables such as teaching functions, reported in preceding chapters, with concrete examples of the meaning of these variables in actual classrooms. The information came from test scores, ratings, log and observer records, attendance records, teacher interviews, and a large set of descriptive information provided by field workers. Both the descriptions of classes and the interpretation of trends in the data are subjective and inferential. No statistical tests were run. Only a small sample of classes were studied. The conclusions that are offered should be considered tentative. They suggest possible applications of the BTES model or possible variables for further study.

From examination of data for all 25 second grade classes it was observed that residualized mean achievement on reading and mathematics correlated approximately .50. The correlation between the measures of reading attitude and mathematics attitude was also approximately .50. The correlation between reading attitude and residualized reading achievement was low positive as was the correlation between mathematics attitude and mean residualized mathematics achievement.

The general association of achievement and attitude appeared to be basically orthogonal. Some factors contribute to increased achievement, other factors promote more positive attitudes.

Class scores on ALT variables and teaching process variables were given, to provide concrete instances of the level of these variables in outlier classes. All of the relationships reported in earlier chapters could be observed in these classes:

1. Positive outliers allocated more time to reading than negative outliers. Negative outliers allocated more time to mathematics than positive outliers (other facets of ALT reduce the quality of this time). Positive outliers had less wait time before an assignment was given;
2. Engagement was higher and more homogenous for the positive outlier classes;
3. Students in the positive outlier classes received a higher proportion of easy tasks (low error rate);
4. Students in positive outlier classes received more substantive, interactive instruction, especially academic feedback;
5. Positive outlier teachers were able to make better diagnoses as reflected by the teacher's ability to make more accurate predictions as to how well their target pupils would do on test items. They also used diagnosis on an ongoing basis to prescribe instruction which was rated appropriate to the needs of individual students.

6. Positive outlier teachers clearly valued and emphasized cognitive learning. For this reason they had below average scores on orientation toward affect. However, the classroom atmosphere was relatively friendly and warm as well as work-oriented. Students were held responsible for working but at the same time they sometimes helped each other.

For any one variable, some exceptions could be found. One negative outlier sometimes looked more like the positive outliers than the other two on a particular dimension. But the positive outliers gave a convincing picture of some specific way to integrate these elements into a high-quality instructional experience.

The description of these classes also brought to light some additional ways in which the positive and negative outliers differed. These additional factors provide hypotheses for further study:

1. Two of the three positive outliers were split-grade classes. One was a combination 1-2; one was a combination 2-3;
2. Positive outliers had a larger number of categories of assistance available to the teacher such as aides, adult volunteers specialists, and/or tutors (possibly a correlate of participation in the ECE program).
3. Most grade 2 classes in the study were on a staggered reading schedule (half the students at school for reading early in the morning, the other half arriving late but staying for reading in the afternoon). Of the few classes that did not have this schedule, two were negative outliers.
4. Positive outlier classes had some system of positive rewards for achievement.
5. Negative outlier classes appeared to show, in a comparative sense, less growth from the A to B testing occasions than they did from the B to C testing occasions. This is to say that the negative outliers appeared to take longer than the other five classes to "get going" academically.

CHAPTER 9: STUDENT ATTITUDES AND SCHOOL INSTRUCTION

In this chapter, relationships between student attitude and the three major sets of variables in the ALT model (teaching processes, ALT, and student achievement) are examined. Student attitudes are not directly represented in the ALT model as it currently stands. The model does not hypothesize the direction or strength of relationships between student attitudes and the constructs in the model. It would be desirable, in future development of the ALT model, to include student attitudes as a component of the model. Toward this goal, it will be necessary to understand the relationship between student attitudes, teaching processes, and academic learning time.

Analyses in this chapter were intended to explore the relationships of student attitudes with teaching and learning. However, no specific relationships were hypothesized. It was not assumed that effective teaching behaviors or student learning behaviors would necessarily show a positive relationship with change in student attitude. Nor was it assumed that they would show a negative relationship with attitude change. Rather, there was simply a general question: Are changes in student attitudes over the school year related to the elements of the ALT model?

There was, however, one specific issue of importance. This is the possibility that ALT may result in more negative student attitudes toward reading, mathematics, and school. There is some concern that the learning behaviors represented in the ALT model may consist, essentially, of hard work in the "sweatshop" tradition. This might result in short-term achievement gains, but a long-term deterioration of academic attitudes and enthusiasm for learning. Therefore, a central purpose of this chapter is to investigate the possible negative impact of ALT on student academic attitudes. We hope, however, not to

find it.

It should be emphasized, here, that change in attitudes is difficult to measure accurately, particularly with younger students (such as second graders). Measurement issues will be discussed, but it will not be possible to provide a definitive evaluation of the reliability and validity of the attitude change scores. Given the complications of measuring attitude change, in addition to the exploratory nature of these analyses, the criteria for interpreting results will be flexible. It will be assumed that measurement errors have diminished the size and consistency of effects that may be found. For exploratory purposes, however, this is not too critical. Having no hypotheses to test, it is not necessary to reach definitive conclusions. Trends in the results will be interpreted wherever those trends show some consistency across quadrants and across related variables. Low and/or nonsignificant correlations will be considered, if they contribute to a larger pattern. Furthermore, variations from a pattern will be ignored, if they appear to be relatively trivial. Consistency across quadrants and related variables will be given more attention than an isolated, significant correlation.

This approach is useful for exploration because it avoids premature dismissal of valid hypotheses that may be inadequately assessed by these data. In addition, this approach increases the likelihood of detecting any detrimental effects of ALT on attitude. It is, of course, important to highlight the speculative nature of interpretations based on loosely defined patterns.

Description of the Attitude Data

Student attitudes toward reading, mathematics, and school were assessed at the A, B, and C testing occasions. At grade 2, students responded on three point Likert-type scales. At grade 5, seven point scales were used. Item

responses were summed to provide the three scale scores at each grade level. An index of change in attitude was obtained by regressing attitude scores at C on attitude scores at A, thereby generating residualized scores on student attitude toward reading, mathematics, and school. These residualized scores were intended to reflect change in student attitude over the school year (between A and C). Residual attitude was correlated with measures of achievement (also residualized), ALT, and teaching process.

Table 9-1 presents the scale ranges, means, standard deviations, and alpha coefficients for the three measures of student attitude at each of the three testing occasions. It is clear that these data show no important overall change in attitude across the school year for second or fifth grade students. There is a slight tendency for attitude to increase over the year at the second grade and to decrease at grade five, but these trends are inconsequential relative to the variance among students on these measures.

Table 9-1 also indicates that the three attitude scales are reasonably reliable in terms of internal consistency (represented by alpha coefficients). The alpha coefficients fell below .70 only for attitude toward school assessed at A, where the coefficients were .58 and .55 for grades 2 and 5 respectively. It is also encouraging that the reliabilities do not appear to be appreciably lower at grade 2 than they are at grade 5. This would suggest that the younger students understood how to respond well enough to do so consistently.

There is some indication of a positive response bias, particularly at grade 2 where the mean scores tend to be about one standard deviation below the highest score possible on each scale. The grade 5 distribution is also positively skewed, although less so (with means that are about two standard deviations below ceiling). This puts a limit on the amount of positive change in attitude that was possible.

Table 9-1

Scale Ranges, Means, Standard Deviations, and Alpha Coefficients
for Student Attitude Toward Reading, Mathematics,
and School at the A, B, and C Testing Occasions

Attitude Variable	Scale Range	Grade 2 (N=139)		
		Testing Occasion		
		A	B	C
Attitude toward Reading	7-21	15.4 (3.6) [.71]	15.8 (4.0) [.82]	16.6 (4.0) [.85]
Attitude toward Mathematics	7-21	16.0 (4.1) [.83]	16.0 (4.4) [.88]	16.9 (4.2) [.89]
Attitude toward School	2-6	4.9 (1.3) [.58]	4.8 (1.5) [.76]	5.0 (1.2) [.71]
Attitude Variable	Scale Range	Grade 5 (N=122)		
		Testing Occasion		
		A	B	C
Attitude toward Reading	11-55	42.3 (7.6) [.83]	40.6 (8.7) [.83]	40.2 (7.8) [.88]
Attitude toward Mathematics	11-55	38.8 (9.8) [.90]	38.6 (10.6) [.93]	37.1 (10.6) [.93]
Attitude toward School	2-10	7.5 (2.2) [.55]	7.5 (2.2) [.73]	7.2 (2.1) [.72]

Entries are: Means (Standard Deviation) [Alpha Coefficients]

It is also possible to make rough estimates of the reliability of residual attitude scores (using the same procedure as that used in Chapter 4, formula number 50 in O'Connor, 1972). These estimates are generally quite high for the reading and mathematics scales (for example, .82 for grade 2 reading attitude residuals and .87 for grade 5 mathematics attitude residuals). However, they are lower for residual attitude toward school (.66 at grade 2 and .52 at grade 5).

It should be cautioned, however, that all estimates of reliability discussed here are based on alpha coefficients of internal consistency. Hence, these coefficients simply indicate that the items within a scale were positively associated with the other items in the same scale. That is, if a student gave a relatively low response to one item on attitude toward reading, then he/she was likely to also give relatively low responses on the other reading attitude items. However, it is particularly possible with attitude scales that students may respond consistently at any given point in time, but respond inconsistently across different points in time. That is, a student may consistently say that he/she likes reading just after having enjoyed a good book, but may respond much less positively the next day after finishing a difficult test on that book.

Not only is it difficult to be sure that these attitude scales are reliable, but it is almost impossible to be sure that they are valid. That is, the scales may be reliable insofar as students respond consistently at any given point in time or perhaps even across points in time. However, it is almost impossible to know with certainty what a given student attitude score really represents. The attitude items essentially just ask the student how much he/she likes reading, mathematics, and school. This could reflect true "liking" or "disliking." It could also reflect what peers or adults have taught the student to express. Various other possibilities could be proposed. It is even possible that some

students respond on the basis on one set of reasons, while other students respond on an entirely different basis. All of these possibilities could yield consistent, or reliable, student "attitude" scores.

Given the difficulties of evaluating the reliability and validity of these attitude scales, interpretations of analyses in this chapter should be considered tentative. Even relatively strong and consistent results for variables such as attitude should be replicated in other studies before they are looked upon as well substantiated.

Table 9-2 presents correlations among the three measures of residual student attitude. Note that all six correlations are positive and that four of them are significant at the .05 level. The three correlations at grade 5, all significant, are somewhat stronger than those at grade 2. The tendency for residual attitude toward reading, mathematics and school to be positively associated is not surprising, in that all three variables involve attitude change related to academic classroom events. It is also reasonable that these correlations are moderate or low (ranging from .11 to .41), because each of the three variables involves a somewhat different domain and each is subject to a certain degree of measurement error.

Student Attitude and Achievement

Table 9-3 presents the correlations of residual achievement in total reading and total mathematics with each of the three residual attitude variables. Only one of the 12 correlations, that between total mathematics and attitude toward school at grade 5, reaches significance ($r=.20$, $p \leq .05$). There is a slight, overall tendency for residual achievement and residual attitude to be positively correlated in each quadrant except grade 2 mathematics (where the correlations are very close to zero). It is possible that improvements in student achievement levels result in improved student attitudes toward academic

Table 9-2

Correlations Among Measures of Student Attitude Toward
Reading, Mathematics, and School
A-C Period

Grade 2 (N=139)					
Variable	Mean	Standard Deviation	Correlations		
			1	2	3
1. Residual Attitude ^a toward Reading	-.01	3.8	1.00		
2. Residual Attitude toward Mathematics	-.01	3.8	.16	1.00	
3. Residual Attitude toward School	0.0	1.2	.40*	.11	1.00
Grade 5 (N=122)					
Variable	Mean	Standard Deviation	Correlations		
			1	2	3
1. Residual Attitude toward Reading	-.04	6.7	1.00		
2. Residual Attitude toward Mathematics	-.01	9.9	.27*	1.00	
3. Residual Attitude toward School	-.01	1.8	.32*	.41*	1.00

^a Residuals were obtained from the "C" measure regressed on the corresponding "A" measure. Non-zero mean residuals are the result of rounding error.

*Starred correlations are significant at the .05 level using student-level degrees of freedom.

Table 9-3

Correlations of Residual Achievement
with Residual Attitude ^a
A-C Period

Residual Achievement Variable	Grade 2 (N=139)			Grade 5 (N=122)		
	Residual Attitude			Residual Attitude		
	Reading	Math	School	Reading	Math	School
Total Reading	.08	.12	.08	.13	.15	.13
Total Mathematics	-.07	.03	-.04	.00	.10	.20 *

^a Residuals were obtained from the "C" measure regressed on the corresponding "A" measure.

* Starred correlations are significant at the .05 level using student-level degrees of freedom.

instruction. These data, however, lend this conclusion only tentative support.

Table 9-4 presents the correlations of achievement retention over the summer (as measured by achievement at D residualized on achievement at C) with residual attitude. Only a few, selected subscales in reading and mathematics were administered at each grade. Therefore, correlations are reported for residual achievement as measured with these individual subscales, rather than total reading or total mathematics. None of the 27 correlations presented in Table 9-4 is significant at the .05 level. Furthermore, there is no consistent trend across the nonsignificant correlations. The majority of the grade 2 correlations are positive (10 of 15), while exactly half of the grade 5 correlations are positive. It may be concluded that no relationship was shown between achievement retention over the summer and attitude change over the preceding school year.

No strong association between attitude change and achievement change is found in this study. There is, however, a slight trend toward a positive association of change in achievement and change in attitude. No trend is noted for associations between summer retention and attitude change over the preceding school year.

Student Attitude and Academic Learning Time

Table 9-5 presents correlations between the ALT variables (time allocation, engagement rate, low error rate, and high error rate) and residual student attitude. No significant correlations (at the .05 level) are shown for either time allocation or engagement rate. Furthermore, the nonsignificant correlations for these variables show about the same number of positive and negative associations with attitude change. Therefore, it may be concluded that no association has been found for either time allocation or engagement rate with student attitude change in reading or mathematics at grade 2 or 5.

Table 9-4

Correlations of Summer Retention
(Residual Achievement, C-D Testing)
with Residual Attitude (A-C Period)

Grade 2 (N=97)			
Residual Achievement ^a Variable (Summer Retention)	Residual Attitude ^b		
	Reading	Mathematics	School
Word Structure Meaningful Units	.05	.10	.15
Word Structure Syllables	-.12	.05	.04
Total Comprehension	.15	.03	.16
Addition and Subtraction Speed Tests	-.08	-.04	.01
Place Value, Numerals	.00	.17	-.02
Grade 5 (N=89)			
Residual Achievement ^a Variable (Summer Retention)	Residual Attitude ^b		
	Reading	Mathematics	School
Total Word Meaning	.11	-.04	.02
Multiplication	-.11	.04	-.08
Division	.15	.05	.02
Fractions	-.16	-.08	-.04

^a Only subtests for those achievement categories listed were administered at testing occasion D. Residuals were obtained from the "D" measure regressed on the corresponding "C" measure.

^b Residuals were obtained from the "C" measure regressed on the corresponding "A" measure.

Note - None of the correlations in this table were significant at the .05 level using student-level degrees of freedom.

Table 9-5

Correlations of ALT Variables with Residual Attitude
A-C Period

ALT Variable	Grade 2 (N=139)			Grade 5 (N=122)		
	Residual Attitude ^a			Residual Attitude ^a		
	Reading	Math	School	Reading	Math	School
Time Allocated to Reading	.04	-.04	.06	.10	-.08	-.10
Time Allocated to Mathematics	-.09	-.06	.01	.00	.01	.00
Engagement Rate during Reading	-.08	-.07	.11	.02	-.10	.08
Engagement Rate during Mathematics	.00	.05	.14	-.17	-.11	.09
Low Error Rate in Reading	.05	.06	.25*	.07	.09	.09
High Error Rate in Reading	-.08	-.16	-.22*	-.10	-.02	.19*
Low Error Rate in Mathematics	.13	.19*	.09	.06	.12	.15
High Error Rate in Mathematics	-.15	-.10	-.12	-.01	.18*	.06

^a Residuals were obtained from the "C" measure regressed on the corresponding "A" measure.

* Starred correlations are significant at the .05 level using student-level degrees of freedom.

Five out of 24 correlations between error rate and attitude change are significant at the .05 level in Table 9-5. Considering both significant and nonsignificant correlations, there is a consistent positive association between low error rate and residual attitude across all four quadrants. However, the results for high error rate are more complex. At grade 2, high error rate shows a consistent negative association with residual attitude. At grade 5, however, three positive and three negative correlations are shown, where two of the positive correlations reach significance. Therefore, the association of high error rate and change in student attitude at grade 5 is unclear, although it tends to be positive more so than negative.

Summarizing the correlations of ALT with residual attitude, time allocation and engagement rate were found to be unassociated with residual attitude, while error rate did show significant associations. Low error rate was positively associated with residual attitude at both grades. High error rate was negatively associated with residual attitude in grade 2. At grade 5 however, high error rate tended to be positively associated with residual attitude, although this trend was inconsistent. Generally, it would appear reasonable that students with a low error rate on instructional tasks would have a more positive attitude toward instruction, while students with a high error rate would have a more negative attitude. The trend toward a positive association between high error rate and attitude change at grade 5 is difficult to understand. High error rate did show a consistently strong negative association with student learning outcomes (see Chapter 4). It is possible, however, that fifth grade students have a greater tolerance for experiencing failure than do second grade students. This might account for the lack of a negative association between high error rate and residual attitude at grade 5.

Student Attitude and Teaching Process Measures

Measures of teaching process variables included both planning or general variables and interactive teaching behavior variables (see Chapter 3 for discussion of these variables). Correlations of planning or general variables with residual attitude are presented in Table 9-6, while correlations of interactive teaching behaviours with residual attitude are presented in Table 9-7. Eight correlations, four per grade, are presented for each of the teaching process variables in Tables 9-6 and 9-7. Most of the teaching process variables were obtained separately for reading and mathematics instruction (variables G1, G2, G3, and G4 are exceptions; see Chapters 5 and 6 for discussion). Residual attitude in reading is correlated with each teaching process variable as measured for reading, while residual attitude in mathematics is correlated with each teaching process variable measured for mathematics. Residual attitude toward school was correlated with each teaching process variable as measured for both reading and mathematics. (Hence there are two columns under the "school" headings in these tables, one for reading and one for mathematics. Variables G1, G2, G3, and G4 have single entries under the "school" headings because these variables were not measured separately for reading and mathematics).

Table 9-6 shows that few of the planning or general teaching process variables had more than one significant correlation out of the eight presented for each of them. (Using the .05 level for significance, one of eight is about what would be expected by chance).¹ Furthermore, none of the 11 teaching process variables is consistently positive or consistently negative in its correlation with residual attitude. There are, however, some interesting trends, particularly at grade 5.

The strongest effects were found for orientation toward affect at grade 5,

¹In addition, significance levels are determined liberally for correlations involving the class-level planning or general variables, as discussed in Chapter 7.

Table 9-6

Correlations of Planning or General Variables
with Residual Attitude
A-C Period

Planning or General Variable	Grade 2 (N=139)				Grade 5 (N=122)			
	Residual Attitude ^a				Residual Attitude ^a			
	Reading	Math	School		Reading	Math	School	
			Reading	Math			Reading	Math
D1 Item Prediction	.07	.14	-.04	-.13	.18*	.08	-.05	.02
D2 Knowledge of Subject Matter	-.04	-.02	-.07	-.10	-.19*	.09	-.23*	-.07
D3 Differentiated Perceptions	.05	.12	.00	-.02	.08	.04	.07	.18*
P1 Goal Orientation	-.05	.02	.16	.14	.15	-.07	-.04	-.10
P2 Program Change- Need	-.13	-.32*	-.02	-.15	.20*	-.11	.15	.17
P3 Flexibility of Curriculum	.04	-.05	-.04	.04	.06	.09	.04	.09
G1 Global Rating (Like Class?)	-.24*	-.17	-.04		-.22*	-.01		.01
A Appropriateness of Curriculum for Individuals	-.10	.10	-.04	.00	.21*	-.01	-.02	.03
G2 Perceived Academic Competence	-.12	.02	-.03		.03	-.12		-.26*
G3 Orientation Toward Affect	-.03	-.05	-.04		.01	.20*		.18*
G4 Learning Environment (Cooperation and Student Responsibility)	-.05	.07	-.13		.09	-.02		-.04

^a Residuals were obtained from the "C" measure regressed on the corresponding "A" measure.

* Starred correlations are significant at the .05 level using student-level degrees of freedom. The use of student-level degrees of freedom is a liberal interpretation of the class-level planning or general variables (see discussion in Chapter 7).

Table 9-7

Correlations of Interactive Teaching Behaviors
with Residual Attitude
A-C Period

Interactive Teaching Behavior Variable	Grade 2 (N=139)				Grade 5 (N=122)			
	Residual Attitude ^a				Residual Attitude ^a			
	Reading	Math	School		Reading	Math	School	
			Reading	Math			Reading	Math
Explanation-Need	-.04	-.08	-.01	-.07	.08	.12	.16	.04
Explanation-Planned	-.10	-.01	.06	.02	-.24*	.07	-.02	.21*
Academic Monitoring	-.04	-.06	.05	.05	-.15	.18*	.07	-.02
Academic Feedback	-.09	.11	.04	.00	-.02	.14	.12	.07
Total Substantive	-.11	-.01	.06	.01	-.13	.18*	.12	.12
Structure/Direct	-.05	.10	.16	.01	.11	.02	.23*	.14
Task Engagement Feedback	-.18*	-.12	-.04	-.09	-.14	.15	.09	.16

^a Residuals were obtained from the "C" measure regressed on the corresponding "A" measure.

* Starred correlations are significant at the .05 level using student-level degrees of freedom.

where two of three correlations with residual attitude were significant and positive. Orientation toward affect did not correlate significantly with residual attitude at grade 2.

There was also an indication that variables reflecting the teacher's tendency to adapt instruction according to individual student needs was positively associated with residual student attitude at grade 5, but not at grade 2. Program change-need shows one significant positive correlation at grade 5 (two out of three nonsignificant correlations are also positive). However, the same variable shows one significant negative correlation at grade 2 (and three nonsignificant negative correlations).²

Three other variables, item prediction, differentiated perceptions, and appropriateness of curriculum for individuals, each show one significant positive correlation with residual attitude at grade 5, but no significant correlations at grade 2. Item prediction and differentiated perceptions are related to the teacher's awareness of the individual characteristics and needs of different students (see Chapter 6). Appropriateness of curriculum reflects the extent to which the difficulty levels and pacing of instruction are adjusted to the needs of individual students (see Chapter 6). Overall, then, these findings show a positive association at grade 5 of the teacher's tendency to recognize individual student needs and to adapt instruction to those needs with the student's change in attitude over the school year. This pattern is not strong for any single variable. However, its reappearance for all four variables involving recognition of individual student needs provides at least tentative indication of the pattern at grade 5.

²It is interesting to note that program change-need is negatively correlated with low error rate only in grade 2 mathematics (see Chapter 7), which is also where it correlates negatively with attitude.

Another trend at grade 5 was the negative association of the teacher's knowledge of subject matter and residual student attitude (two significant correlations out of four). Perceived academic competence was also negatively correlated with residual student attitude at grade 5 (one significant correlation out of three). Neither variable was significantly correlated with student attitude change at grade 2.

One last, and somewhat puzzling, trend is the negative association of the observer's global rating of the class and residual student attitude at both grades (one significant correlation out of three at each grade, both for attitude toward reading). It is reasonable to conclude that the observers reacted positively to different classes than did the students in this study. The observers appear to have "liked" different aspects of instruction than did the students. It is difficult to explain why observers and students differed, but well worth being aware that they did.

It may be possible, with some interpretive license, to detect a general pattern in the relationships of teacher planning and general process variables with student attitude change at grade 5. The findings presented in Table 9-6 may indicate that grade 5 students react positively when their teachers seem to give them individual attention. Fifth grade student attitude change is positively correlated with variables involving the teacher's affective and academic recognition of their individual needs and differences. This is indicated by the positive correlations of residual attitude at grade 5 with orientation toward affect, program change-need, item prediction, differentiated preceptions, and appropriateness of curriculum for individuals. However, grade 5 students may react negatively (in attitude) toward the teacher's concern for substantive academic curriculum per se, as distinguished from the teacher's concern for the relationship between the academic curriculum and student needs. This would be

indicated by the negative correlations of fifth grade student attitude change with teacher academic knowledge and competence. It is possible that fifth grade students react negatively to teacher academic knowledge and competence when it reflects a greater teacher concern for the substance of instruction than for the relationship of the instruction to individual students. This could, then, be related to a negative student reaction to the lack of individual attention from the teacher. Bear in mind, however, that these are tentative conclusions based on liberal interpretations of the data.

Table 9-7 presents correlations of interactive teaching behaviors with residual student attitude. Few significant correlations (at the .05 level) and no consistent trends are indicated for the substantive teaching behaviors (the first five interactive teaching variables). The only significant correlations for substantive teaching behaviors occur at grade 5. Two of these are correlations of explanation-planned with grade 5 residual student attitude, but one is negative while the other is positive. Academic monitoring shows a significant positive correlation with grade 5 residual attitude, but this variable also shows a nonsignificant negative correlation that is nearly as large as the significant positive coefficient. The data may show a tendency for total substantive behaviors to be positively correlated with residual attitude at grade 5, where three coefficients are positive and one of these is significant. However, the fourth coefficient is negative and of a magnitude comparable to that of the other three.

There is some indication of moderate associations of non-substantive teaching behaviors with residual student attitude. Structuring and directing behaviors show a positive, significant correlation at grade 5 that is supported by consistently positive, nonsignificant correlations at that grade. Task engagement feedback shows a negative, significant correlation at grade 2 that is also supported by consistently negative, nonsignificant correlations. This may in-

dicade that grade 5 students respond positively to structure provided by the teacher, while grade 2 students respond negatively to disciplinary feedback from the teacher. However, these conclusions receive only weak support from the data and should be treated as tentative.

Summary

No strong and consistent patterns were found for relationships of residual student attitude with student achievement, student ALT, or teaching process variables. However, some trends were indicated that warrant cautious consideration.

Residual student attitude tended to show low, positive associations with residual student achievement, although this trend was not always consistent. Analyses of summer retention showed no relationship with residual student attitude for the preceding school year.

Analyses of ALT showed no relationships of time allocation or engagement rate with residual student attitude. However, low error rate showed a consistent though low positive association with residual attitude. High error rate was negatively associated with residual attitude at grade 2, but not at grade 5. In fact, an inconsistent positive association of high error rate with residual attitude appeared at grade 5.

Analyses of the planning or general teaching process variables showed some interesting, but tentative, trends at grade 5. Residual attitude at grade 5 was positively correlated with orientation toward affect, program change-need, item prediction, differentiated perceptions, and appropriateness of curriculum for individuals. Residual attitude at grade 5 was negatively associated with knowledge of subject matter and perceived academic competence. A puzzling negative correlation of residual attitude with the observer's global rating of the class was found at both grades. In addition, residual attitude at grade 2 was negatively correlated with program change-need, in contrast with the positive association

of these variables at grade 5.

These findings for planning and general teaching process variables at grade 5 may be liberally interpreted as showing a pattern of support for an association of teacher attention to individual students with positive changes in student attitude over the school year. It is possible that students get less individual attention as they progress to higher grade-levels. This, in turn, could lead to a heightened student sensitivity to the degree of individual attention that they do receive. The speculation in this conclusion, however, should be underscored.

Analyses of interactive teaching behaviors showed no consistent associations of substantive teaching behaviors with residual student attitude. There was some indication of moderate associations of non-substantive teaching behaviours with residual student attitude. Structuring and directing behaviors were positively correlated with residual attitude at grade 5, while task engagement feedback was negatively correlated with residual attitude at grade 2.

Overall, it may be concluded that these data show a surprisingly low, though positive, association of change in student achievement with change in student attitude. There is also an encouraging lack of association of time allocation and engagement rate with student attitude change. This indicates that relatively high levels of time allocation and engagement are not generally associated with aversive, "nose-to-the-grindstone" classroom environments. Furthermore, improved student attitude tended to be positively associated with the error rate component of ALT, as indicated by the low, positive association of low error rate with residual student attitude. Regarding teacher process variables, it may tentatively be concluded that improvement in student attitude was positively associated with individual attention from the teacher at grade 5, higher structuring at grade 5, and less disciplinary feedback at grade 2.

It should be restated that the reliability and validity of the attitude measures are difficult to evaluate. In addition, no strong and consistent patterns emerged from analyses of these measures. The most important finding of these analyses was the lack of any indication that higher levels of ALT were associated with declining student attitudes.

CHAPTER 10. SUMMER RETENTION OF ACHIEVEMENT

Introduction

Earlier chapters examined student achievement gain over the school year and presented analyses relating these measures to teaching process variables and academic learning time. The primary consideration in those analyses was the identification of teaching behaviors and student learning behaviors which are associated with measured learning gains within one school year. A number of other considerations warrant our attention. For example, do students retain the knowledge gained over the school year when tested at some later time? Are there some characteristics of instruction which tend to foster the retention of knowledge more than other characteristics? Do those students who gain the most over the school year tend to maintain their gains over periods of no instruction or are large gains in the school year associated with relatively large losses over the ensuing vacation period? This general area is important to education since we all expect that the effects of schooling will be both cumulative and retained over time.

In Phase III-B of BTES, we addressed two questions in the area of achievement retention. First, what is the pattern of student achievement scores when considered over a period with regular school instruction followed by a period with no instruction? Second, what characteristics of instruction during the school year are related to subsequent retention of achievement?

To address these questions, students in the Phase III-B sample were retested in early September, 1977. This administration was referred to as the D testing. Thus, test scores were available for the A, B, C, and

D occasions in addition to the instructional process data collected during the A-C interval. The first question about the pattern of achievement retention is addressed by examining the profile of scores for the same students on the same subscales at these four points in time. The second question, about what instructional process measures during the A-C period are related to achievement at D, is addressed by a series of regression analyses and other correlational procedures.

Sample

When the D testing took place, the students in the study were just beginning a new school year. Hence, most students were in new classes with different teachers, and in many cases, students were in new schools. As a result of this mobility, not all students in the total Phase III-B sample were located and tested at the D occasion. The student samples available for analyses of achievement retention at both grades were subsets of the total Phase III-B sample. For grade 2, 97 of 139 students were tested at D while for grade 5, 89 of 122 were tested.

The students who were not tested at D may have had some common characteristics which might cause their achievement retention pattern to be different from the pattern displayed by those who were tested. If this were true then analyses of the available subsample would not be representative of the Phase III-B sample as a whole. To check this possibility, we compared scores at C for those students who were tested at D with scores for students who were lost to the study by attrition. Table 10-1 presents means and standard deviations for these comparison groups on fourteen subtests. The right hand column shows the probability of obtaining a difference between the group means by chance alone. Thirteen of

Table 10-1

Comparison of Achievement Scores at C Between Students Who Were Tested at D
and Students Who Were Not Tested at D

Subtest	Grade	Students Who Were Tested at D			Students Who Were <u>Not</u> Tested at D			Probability ¹
		N	Mean	S.D.	N	Mean	S.D.	
Word Structure - Meaningful Units	2	97	17.7	4.8	42	17.1	5.5	0.49
Word Structure - Syllables	2	97	5.1	2.9	42	6.0	2.6	0.10
Total Comprehension	2	97	11.1	6.9	42	12.4	7.3	0.34
Total Reading	2	97	90.2	28.3	42	91.9	35.3	0.77
Addition and Subtraction- No Regrouping	2	97	25.7	6.4	42	25.7	7.2	0.99
Addition and Subtraction- Speed Test	2	97	17.3	6.5	42	15.3	7.7	0.13
Place Value/Numerals	2	97	12.8	6.4	42	13.5	6.6	0.56
Total Mathematics	2	97	95.4	29.6	42	93.7	32.8	0.76
Total Word Meaning	5	89	29.4	11.5	33	27.2	14.8	0.39
Total Reading	5	89	73.0	27.4	33	67.3	33.1	0.34
Multiplication	5	89	10.8	4.7	33	12.8	3.4	0.01
Division	5	89	7.2	5.6	33	7.8	5.4	0.62
Fractions	5	89	5.4	4.9	33	6.2	4.9	0.46
Total Mathematics	5	89	86.1	30.6	33	92.0	18.9	0.21

¹This is the probability that the means of the two groups are equal.

fourteen comparisons showed that the mean achievement at C for the group that was tested at D was not different at the .05 level of significance from the mean at C for the students who were "lost" over the summer.

Based on this evidence, the group which was tested at D did not differ from the students lost by attrition. We proceeded with the examination of the two questions about retention of achievement, concluding that results based on the D sample would be generally representative of the whole Phase III-B sample. All analyses reported in this chapter are based on the D sample (97 students at grade 2, and 89 students at grade 5).

Pattern of Achievement Scores A-D

The pattern of achievement scores for the D sample is described by the means and variances for the subtests at the four testing occasions. Before examining the data, two points need to be discussed. First, the time between testing occasions was not constant. Between the A and B testing occasions approximately seven weeks elapsed. The number of weeks between B and C, and between C and D were approximately 19 and 18 respectively. These differences in the lengths of the intertest period should be kept in mind when considering the achievement patterns. That is, gains in achievement from B to C are expected to be considerably greater than gains from A to B.

The second point concerns the amounts of instruction which took place in the intertest periods. The A-B period was all regular school instruction. The B-C period of approximately 19 weeks included two weeks of Christmas vacation and one week of Spring vacation for almost all students. The remaining 16 weeks was regular school instruction. For the school instruction during the A-B and B-C periods we have collected large amounts of instructional process data. The C-D period, which is of most interest

in this chapter, is somewhat complicated. For example, school ended in June for most students but the C testing occurred in early May. Therefore, most students in the sample had about six weeks of regular school instruction during the C-D period for which we have no specific process information. In addition, the time between June and the D testing in September included differing amounts of instruction for the students. Some students attended summer schools while others did not.

We were able to identify all students who received academic instruction during the summer although we do not know what content was covered or other characteristics of the instruction. Hence, at grade 2 we divided the sample into two subgroups: one that received no summer instruction and one that did receive academic instruction during the summer. At grade 5, several students attended year round schools. In these schools, students spend about the same number of days in school instruction as students in regular schools but instruction is arranged in 10 week cycles. Students have about three weeks off between cycles. Hence several students at grade 5 had regular schooling during most of the time when summer vacation ordinarily takes place. In this case, the sample was separated into three groups: one group that had no academic summer instruction, one group that attended year round schools, and one group that had summer academic instruction in some other context.

Means and standard deviations on a total of nine subtests are shown for the four testing occasions in Tables 10-2 and 10-3. In each case the data are broken out by subgroups representing different amounts of summer instruction.

For grade 2, both groups showed an increasing trend on achievement

Table 10-2

Mean Achievement on Four Testing Occasions
Grade 2

Variable	Group	Number of Students	Test Occasion			
			A	B	C	D
Word Structure (Meaningful Units)	No Academic Summer Instr.	81	9.7 (5.2)	12.8 (5.5)	18.0 (4.6)	18.4 (4.3)
	Academic Summer Instr.	16	11.0 (6.2)	13.7 (5.9)	16.3 (5.8)	18.2 (4.4)
	Total	97	9.9 (5.3)	13.0 (5.5)	17.7 (4.8)	18.4 (4.3)
Word Structure (Syllables)	No Academic Summer Instr.	81	2.3 (2.4)	2.7 (2.5)	5.1 (2.9)	5.4 (2.9)
	Academic Summer Instr.	16	2.7 (3.5)	1.8 (3.0)	4.9 (3.0)	5.6 (3.1)
	Total	97	2.4 (2.6)	2.6 (2.6)	5.1 (2.9)	5.4 (2.9)
Total Comprehension	No Academic Summer Instr.	81	*	6.5 (6.3)	11.1 (7.0)	12.4 (7.5)
	Academic Summer Instr.	16	*	7.9 (6.3)	11.0 (6.7)	11.4 (7.8)
	Total	97	*	6.7 (6.3)	11.1 (6.9)	12.2 (7.5)
Addition and Subtraction Speed Test	No Academic Summer Instr.	81	8.9 (4.8)	10.9 (6.0)	16.8 (6.4)	17.0 (6.5)
	Academic Summer Instr.	16	8.6 (6.6)	13.5 (6.8)	19.4 (6.9)	16.8 (7.4)
	Total	97	8.8 (5.1)	11.3 (6.1)	17.3 (6.5)	17.0 (6.6)
Place Value	No Academic Summer Instr.	81	4.8 (4.4)	7.3 (5.0)	12.8 (6.3)	14.7 (7.0)
	Academic Summer Instr.	16	5.3 (5.1)	7.7 (6.1)	12.9 (6.9)	14.8 (4.2)
	Total	97	4.9 (4.5)	7.4 (5.1)	12.8 (6.4)	14.7 (6.6)

Entries are means with standard deviations in parentheses.

*This subtest was not administered at occasion A.

Table 10-3

Mean Achievement on Four Testing Occasions
Grade 5

Variable	Group	Number of Students	Test Occasion			
			A	B	C	D
Total Word Meaning	No Academic Summer Instr.	60	20.7 (11.0)	23.4 (11.9)	29.3 (12.0)	28.9 (13.5)
	Academic Summer Instr.	14	18.9 (10.1)	26.4 (13.7)	31.0 (10.5)	28.3 (15.5)
	Year Round School	15	19.5 (10.0)	22.8 (10.7)	28.7 (10.6)	29.2 (13.6)
	Total	89	20.2 (10.6)	23.8 (11.9)	29.4 (11.5)	28.8 (13.7)
Multiplication	No Academic Summer Instr.	60	7.7 (5.8)	9.7 (5.7)	10.8 (4.7)	10.3 (5.1)
	Academic Summer Instr.	14	4.2 (4.8)	6.4 (5.5)	9.1 (5.2)	10.4 (5.5)
	Year Round School	15	9.2 (4.7)	9.7 (5.8)	12.1 (4.0)	13.1 (2.2)
	Total	89	7.4 (5.6)	9.2 (5.7)	10.8 (4.7)	10.7 (4.9)
Division	No Academic Summer Instr.	60	*	5.4 (5.6)	7.3 (5.9)	7.5 (6.0)
	Academic Summer Instr.	14	*	3.0 (3.5)	6.1 (4.9)	7.7 (5.6)
	Year Round School	15	*	5.8 (4.9)	8.0 (5.4)	9.1 (5.0)
	Total	89	*	5.1 (5.2)	7.2 (5.6)	7.8 (5.8)
Fractions	No Academic Summer Instr.	60	*	2.0 (3.8)	5.7 (4.8)	4.9 (5.3)
	Academic Summer Instr.	14	*	0.7 (3.3)	3.9 (4.8)	4.4 (3.8)
	Year Round School	15	*	2.7 (4.2)	5.7 (5.4)	6.0 (5.2)
	Total	90	*	1.9 (3.8)	5.4 (4.9)	5.0 (5.0)

Entries are means with standard deviations in parentheses.

*This subtest was not administered at occasion A.

across the four testing occasions. In 26 of 28 cases there were positive achievement gains over testing periods. As expected, the gains over the B-C period were largest reflecting the relatively long time period and the presence of school instruction.

At grade 5, positive achievement gains occurred on 26 of 30 occasions (see Table 10-3). All four cases of achievement drops occurred over the C-D period and three of the drops were associated with the no summer instruction group. In no case was the drop substantial.

The overall pattern displayed in the data shows that achievement scores increased over the school year and either increased or stayed about the same over the summer. At grade 5, the year-round school group showed steady increases and at the D testing had somewhat higher scores than the other groups. Substantial summer losses do not appear in these data.

Retention of Achievement C-D

The objective of this section is to identify variables which are related to achievement at D. In accounting for student achievement at D several factors are expected to be important. First the status of students at the C testing must be considered. In addition, some students received summer instruction. Therefore, an index of summer instruction should be included as a potential predictor of achievement at D. After C test scores and summer instruction have been accounted for, what impact do teaching process variables and academic learning time during the school year have on achievement measured at D. Do differences in the way instruction was carried out from October to May have an effect on achievement measured in September? Can an effect of instructional process variables be detected above and beyond the effect of May achievement scores and summer instruction? To shed light on these questions we conducted two kinds of analyses:

a set of regressions and subsequently, correlations between instructional process variables and residual achievement.

First, consider the regression analyses. The purpose of these analyses was to examine the effects of teaching process variables and academic learning time variables assessed over the school year (A-C) on student achievement several months later (at D). Student scores at D on a subtest in reading or mathematics constituted the dependent variable. The independent or predictor variables were student achievement on the identical items at C, a dichotomous index of summer instruction, and several variables representing teaching process and academic learning time over the A-C period. The academic learning time variables included allocated time (for the matching content area), engagement rate, low error rate and high error rate. The teaching process variables were item prediction (D1), appropriateness of curriculum for individuals (A), academic feedback (I4), orientation toward affect (G3), and learning environment (G4). These nine variables representing teaching processes and academic learning time are the same as those included in Chapter 7.

As a first step, we predicted D scores from C scores. Subsequently we predicted D scores from C scores and A scores; and finally we predicted D scores from C scores, summer instruction and the academic learning time and teaching process variables. Table 10-4 shows the proportion of variation in D scores which can be accounted for by the three sets of predictors. As expected, achievement at C is highly related to achievement at D and accounts for from 40 to 70 percent of the variance at D (word structure - syllables was an exception). Hence the C scores are relatively good predictors of achievement at D.

Table 10-4

Achievement Scores at D Predicted by Three Different Sets of Independent Variables
(Grade 2, N=97; Grade 5, N=89)

Subtest	Grade	Percent of Variance at D Accounted for by		
		C Scores	A Scores and C Scores	C Scores, Summer Instruction, ALT, and Teaching Process Variables
Word Structure - Meaningful Units	2	49	52	56
Word Structure - Syllables	2	15	18	38
Total Comprehension	2	58	59	63
Addition and Subtraction Speed Test	2	53	57	59
Place Value/Numerals	2	58	59	66
Total Word Meaning	5	71	76	77
Multiplication	5	40	45	53
Division	5	49	50	60
Fractions	5	64	66	68

The proportion of D variance predicted from A scores and C scores together is also shown in Table 10.4. The contribution of the A scores over and above the contribution of the C scores is indicated by the difference in the percentages for the two sets of predictors. On the average the inclusion of the A scores accounts for an additional three percent of the variance in D scores. The A scores do not add much beyond what is predictable from the C scores alone. This is not a surprising finding. As predictors of D scores, A scores are more remote than C scores. After the A scores were obtained, instruction occurred, and then the C scores were assessed. The C scores reflect both the effects of initial status (the A and C scores are positively correlated) as well as the effects of the intervening instruction. Hence the C scores should predict the D scores better than the A scores predict D scores.

This point can be illustrated by comparing the variance in D scores accounted for by C scores (Table 10-4) with the variance in C scores accounted for by A scores (see Appendix 10-A). C scores account for an average of 47% (grade 2) and 56% (grade 5) of the variance in achievement at D while A scores account for 16% (grade 2) and 24% (grade 5) of the variance in C scores. This is partly because the C and D testing occasions are closer together than the A and C occasions. However, it also reflects the effects of instruction during the A-C period. The fact that some students learned more than others during the school year (as the result of differentially effective instruction), tends to decrease the correlation between A and C achievement scores. During the summer, relatively little instruction took place and the correlation between achievement at C and D was considerably higher.

The final column in Table 10-4 shows the D score variance accounted for by C scores, summer instruction, academic learning time and teaching process variables. By subtracting the percentages in this column from the percentage for the C scores alone, an estimate of the contribution of the teaching process and learning time variables can be obtained. In this case the process and time variables add an average of nine percent to the D variance accounted for in comparison to less than three percent added by the A scores. In earlier chapters we have shown that teaching processes and academic learning time are substantial contributors to student achievement at C. Therefore achievement at C would be expected to carry most of the impact of instruction prior to C when predicting achievement at D. In spite of this expectation, process measures from the A-C period explain a substantial proportion of variance in D scores above and beyond that attributable to C scores (i.e., teaching process and academic learning time measures explain differential retention).

Since summer instruction was included in the set of predictors, it is possible that the additional D score variance was accounted for by summer instruction and not A-C instructional processes. In order to examine the effects of summer instruction and to understand how instructional processes influence retention we conducted another series of regression analyses. In this analysis, predictors were added in a step-wise fashion so that estimates of the proportion of variance in D accounted for by each set could be obtained. Four sets of predictors were entered: C test scores, summer instruction, academic learning time variables, and teaching process variables. Table 10-5 presents results of this analysis for grade 2. The first column shows the proportion of

Table 10-5.

Achievement Scores at D Regressed on Achievement at C, Summer Instruction, Academic Learning Time Variables (AC), and Teaching Process Variables (AC) for Grade 2. (N=97)

Subtest	Percent of Total Variance at D Accounted for by C Scores	Percent of Residual ¹ Variance Accounted for			
		Summer Instruction	Academic Learning Time Variables	Teaching Behavior Variables	Total
Word Structure - Meaningful Units	49	1	4	8	13
Word Structure - Syllables	15	0	22	4	26
Total Comprehension	58	0	10	1	12
Addition and Subtraction Speed Test	53	3	1	10	13
Place Value/Numerals	58	0	12	7	19

¹The residual D score variance is obtained by subtracting the variance accounted for by the pretest (C scores) from the total variance in D scores.

variance in D scores accounted for by C scores. This column is repeated in Table 10-4. The remaining columns estimate the proportion of variance in D, after C scores are accounted for, that is attributable to summer instruction, academic learning time, and teaching process variables when entered in that order. This means that variance in D scores shared by two predictors is attributed to the predictor which is entered first. For example, variance in D scores shared by summer instruction and the academic learning time variables and by summer instruction and teaching process variables is attributed to summer instruction since it is entered earlier in the sequence. Similarly, the academic learning time variables "carry" the variance in D scores shared by academic learning time variables and teaching process variables. The four right-most columns in the table then, indicate proportions of residual variance accounted for by the four sets of independent variables respectively.

For grade 2, summer instruction accounted for residual variance at D in two of the five subtests. The two sets of A-C process measures accounted for substantial proportions of variance above and beyond that accounted for by summer instruction. Hence, academic learning time and teaching process variables are related to achievement at D. Part of their effect is carried by the C scores. However, an additional relatively large effect is also present and accounts for 16% (on the average) of the residual variance in D scores.

Results of a similar analysis for four subtests at grade 5 are presented in Table 10-6. The effect of summer instruction on mathematics achievement is relatively strong. This is very probably due to the inclusion of the year-round schools in the summer instruction group. Otherwise the conclusions drawn for retention of achievement at grade 2

Table 10-6

Achievement Scores at D Regressed on Achievement at C, Summer Instruction, Academic Learning Time Variables (AC), and Teaching Process Variables (AC) for Grade 5. (N=89)

Subtest	Percent of Total Variance at D Accounted for by C Scores	Percent of Residual ¹ Variance Accounted for by			
		Summer Instruction	Academic Learning Time Variables	Teaching Process Variables	Total
Total Word Meaning	71	0	7	14	22
Multiplication	40	4	7	11	22
Division	49	1	14	5	21
Fractions	64	3	5	4	12

10-15

¹The residual D score variance is obtained by subtracting the variance accounted for by the pretest (C scores) from the total variance in D scores.

also fit the grade 5 data. The A-C process variables account for an average of 18% of the residual variance in achievement at D.

The analyses reported in Tables 10-5 and 10-6 correspond in format (with the exception of summer instruction) to analyses run on the total Phase III-B sample using academic learning time and teaching process behaviors to predict C scores (see Chapter 7). By comparison with the earlier analyses, one can determine whether the same types of factors that are associated with greater learning at point C are also associated with greater retention from C to D.¹

Relationship between Instructional Variables and Achievement Retention

The set of instructional process variables covering the A-C period are clearly related to residual achievement at D. We now examine the relationship between specific process variables and achievement. Residual achievement scores (D scores regressed on C scores) were calculated for each student. These scores represent achievement at D after differences in achievement at C have been taken into account. The residuals constitute a retention score. Each of the academic learning time variables and the teaching process variables were correlated with the retention scores. These correlations are shown in Table 10-7. Since summer instruction was related to achievement at D, it is necessary to consider the relationship between summer instruction and both the residual achievement and the instructional process variables when attempting to interpret process

¹Because analyses using D scores were run with a subset of the sample, and included summer instruction, one could question this comparison. Differences could be due to real differences in the factors that influence initial learning versus retention or simply to sampling fluctuations. A-C analyses were rerun on the subsample and the results are shown in Appendix 10-A. In general, patterns were similar for the subsample and the total sample. Those wishing to compare A-C and C-D results might prefer to examine the subsample data.

Table 10-7

Correlations of Academic Learning Time Variables (AC) and Teaching Process Variables (AC)
with Residual Achievement at D

Variable	Summer Instruction				Residual Achievement at D								
	Grade 2		Grade 5		Grade 2 (N=97)					Grade 5 (N=89)			
	Reading	Math	Reading	Math	Word Structure-Meaningful Units	Word Structure-Syllables	Total Comprehension	Speed Test	Place Value Numerals	Total Word Meaning	Multiplication	Division	Fractions
Summer Instruction					.10	.05	-.07	-.18	.00	-.04	.20	.12	.16
Allocated Time					-.02	.32	.21	.04	-.03	.01	-.25	.00	-.03
Engagement Rate	.17	.10	-.17	-.25	.11	.18	.07	.03	.27	-.13	.02	-.02	.14
Low Error Rate	.08	-.07	-.19	-.26	.17	.32	.20	-.03	.24	.19	-.09	.32	.09
High Error Rate	-.10	.00	-.29	.11	-.14	-.12	-.18	-.01	-.21	.11	-.11	-.10	-.03
D1 Item Prediction	.00	-.01	.18	-.24	.11	.20	-.02	.08	.12	.14	.16	-.05	.03
A Appropriateness of Curriculum for Individuals	.15	-.08	.57	.43	.23	.04	.03	.07	.10	.23	.05	-.01	-.06
I4 Academic Feedback	-.22	.10	-.39	-.34	.13	.12	.08	.22	-.05	-.15	.06	.07	.11
G3 Orientation Toward Affect	.12		.29		-.03	.07	-.08	-.06	.03	.25	.11	.14	-.08
G4 Learning Environment (Cooperation, Student Responsibility)	-.01		.16		.02	-.01	.02	-.14	-.03	.02	-.15	-.20	-.04

10-17

NOTE: Correlations of 0.17 (grade 2) and 0.18 (grade 5) are significant at the 0.10 level using the number of students to determine the degrees of freedom.

variable-residual achievement correlations. The first four columns of Table 10-7 report correlations between the summer instruction variable (a dichotomy) and the A-C process measures. For most variables such as engagement rate, four coefficients are shown. Since each student has a reading engagement rate and a mathematics engagement rate, two coefficients can be calculated at each grade level. The correlations indicate that summer instruction was related to A-C process variables for grade 5 but was not consistently related to these variables at grade 2. This adds complexity to the interpretation of the correlations between the process variables and residual achievement especially at grade 5.

If degrees of freedom are based on the number of students, then a correlation of 0.17 (grade 2) or 0.18 (grade 5) is significant at the 0.10 level. A correlation of 0.20 is significant at the 0.05 level with 90 degrees of freedom. Since several of the teaching process variables (A, G3, and G4) are between class variables, use of the number of students to determine the degrees of freedom provides a somewhat liberal interpretation of the correlations (see Chapter 7 for further discussion of this point).

Allocated time during the school year was positively related to residual achievement at D in two of five scales at grade 2. At grade 5 only one substantial correlation (for multiplication) appeared and it was negative. This relationship is very probably due to the year-round school group. From Table 10-3, it is clear that the year-round school students were superior in multiplication. It is likely that some of their mathematics instruction during the C-D period was focused on multiplication skills. The group with no summer instruction had a slight loss from C to D on multiplication. Combining this information with the pattern of achievement of the two groups over the A-C period, we see that both groups had about a three point gain over the A-C period; however, the no

summer instruction group had a slight loss over the C-D period while the year-round school group continued to gain.

Engagement rate during A-C was correlated significantly with residual achievement for two scales. Hence there is a tendency for students with higher attention rates during the school year to sustain that achievement over the summer.

Low error rate was significantly related to residual achievement on six of the nine scales. In each case, the sign of the correlation was positive. The greater the proportion of time students spent working on tasks with low error rate during the school year, the higher their retention scores over the summer period. As expected, high error rate operated in the opposite direction. That is, the smaller the proportion of time students spent on tasks with high error rate, the greater their residual achievement over the C-D period.

Appropriateness of curriculum for individuals (A) correlated positively with residual achievement at D on two reading subtests; one at each grade level. This illustrates that the better the match between curriculum and student needs during the A-C period the higher the retention of achievement from C to D. Each of the other instructional process variables correlated significantly with residual achievement on one subscale.

Summary

Student achievement on the same items was examined at the A, B, C, and D testing periods. In general, achievement levels increased or stayed the same with each test administration. Where students received academic instruction during the summer, scores tended to increase. For students at grade 5 who did not receive academic instruction during the summer,

scores tended to increase from A to B to C. For this group, C and D scores were about the same. These data do not reveal losses, as a general trend, over the C-D period. Since the C testing occasion occurred in early May, approximately five to six weeks of school instruction were included in the C-D period. This fact may have prevented detection of summer losses.

Regression analysis showed that variation in student achievement at D was related to summer instruction. These analyses also showed that academic learning time and teaching process variables based on the school year were related to student achievement at D. These process variables showed an effect on D achievement above and beyond the effect which is shared with the C achievement scores. That is, academic learning time and teaching process variables contributed to achievement at D by contributing to learning during the school year (achievement at C). In addition, academic learning time and teaching process variables have an effect on retention of school year gains (achievement at D) which is independent of C scores.

Residual achievement scores were generated for each student by adjusting achievement at D for differences in achievement at C. These measures of retention were correlated with the academic learning time variables and the set of teaching process variables. Individual variables, especially low error rate, showed consistent relationships with residual achievement.

CHAPTER 11: SUMMARY AND CONCLUSIONS

In Phase III-B of the Beginning Teacher Evaluation Study, hypothesized relationships among teaching process variables, academic learning time variables and student achievement were empirically examined. As a result of previous work on the BTES and other research in the field, an overall framework or model was developed to explain how students learn in elementary school classroom settings. The central construct in the model is Academic Learning Time (ALT), which is defined as the time a student spends engaged on a task with low error rate where the task is directly relevant to a specified set of learning objectives. ALT is intended to measure ongoing student learning through observable classroom behavior. ALT can be partially validated as a measure of learning by examining its association with another measure of learning -- student performance on achievement tests. As a measure of learning, ALT should reflect the influence of instructional treatment variables. Therefore, the academic learning time model of classroom instruction states that students who accumulate more academic learning time in a particular content area will attain higher achievement in that area. The model also states that teaching processes, conceptualized in terms of five teaching functions and classroom context, have a direct influence on academic learning time. According to the model, teaching processes affect student classroom learning (ALT) which, in turn, affects student achievement.

During Phase III-B of BTES, extensive measures of teaching processes, academic learning time and student achievement were systematically collected throughout a full school year. Data were analyzed following a process-product design. The empirical work was carried out in natural classrooms at grades 2 and 5 and focused on acquisition of reading and mathematics skills.

Earlier chapters¹ in this report present the major analyses and findings. This chapter discusses factors to be considered when interpreting the study, presents major conclusions with a summary of supporting results, and discusses the constructs in the ALT model in light of the findings of the study.

What is a "Conclusion?"

All research is a combination of objective analysis and subjective interpretation. Although the image of science is that it is objective, in fact it is very difficult to separate the objective from the subjective. Whenever words are used to describe a phenomenon, subjective interpretation enters in to some degree. Interpretation can play an even stronger part in the identification of causal relationships among variables. As human beings, we are all experts at interpreting complex phenomena, and the concept of causation is basic to our understanding. Researchers try to use systematic procedures in order to make the process of understanding more objective and less prone to error. But they also rely on their reasoning skills and their experience of how the world works.

The role of subjective interpretation differs depending on the way the data were collected. The data analyzed in this study describe naturally occurring events in field-based classrooms. As a result, data analyses are necessarily correlative, where association, but not causation, can be

¹ The material in all sections of this report draws heavily on a series of technical notes and reports generated in earlier portions of Phase III of BTES. An annotated bibliography of these documents is included as Appendix 1-A. Items in Appendix 1-A are grouped according to the three subdivisions of Phase III (Phase III-A, Phase III-A Continuation, and Phase III-B).

demonstrated. To understand the findings of the study, it is essential to recognize the constraints that apply to the interpretation of correlational analyses.

Most research is directed, ultimately, toward a better understanding of causal relationships. It is generally agreed that the experimental research paradigm provides the least ambiguous evidence of causal relationships. Unfortunately, the experimental paradigm has a number of weaknesses when used in studies of teaching and learning. Most importantly, experimental results often fail to generalize to "real world" situations. Also, experiments may overlook some of the best (or worst) treatments that occur in the natural variation among classes. These disadvantages in the use of the experimental paradigm are overcome, to some extent, by studying naturally-occurring instruction in regular classrooms. As a field-based correlational study, BTES examined a wide range of naturally-occurring instructional practices and, hence, should have reasonable ecological validity for existing elementary school classrooms. However, ecological validity is gained at the expense of clearly-defined variables and simplified causal relationships.

This chapter is divided into two basic parts -- "conclusions" and "discussion." A conclusion, in this study, is intended to be relatively objective and close to the data. It tends to be of the type "A is associated with B." Such a statement is an inference from the data that a generalizable, replicable association exists between two variables. It specifically does not state causality. Following each conclusion is a summary of the results on which it is based.

It should be noted that it is not a thoroughly straightforward process to identify conclusions of this type. Some conclusions are identified as less straightforward than others. Not every finding in the study is represented as a conclusion. A relatively large number of potentially

important findings are omitted. Specific findings refer to individual cases of statistical associations between measured variables. In some instances specific findings are contradictory. In other cases, specific findings lend support to a more general statement. The study was designed with a number of types of replication built into it. We were able to empirically evaluate the relationship between two variables within each of two subject areas; within each of two grade levels, and within each of two time periods. The degree of consistency among findings over these quasi-replications was a major factor in identifying sets of findings which could be stated more generally as a conclusion. In some cases, differences between grade levels, subject matters, or time periods are suggested by the data. We did not feel that the data were strong enough to justify major conclusions about such differences, although we often point them out and discuss them in the report. There are a number of conclusions which seem to be generally applicable. Further research is needed before conclusions can be made about more specific or limited relationships.

The reader should also note another complexity in the conclusions. It is not straightforward to define the meaning of all the variables. In a correlational study, no single variable escapes confounding with a multitude of other variables. In addition, operational definitions imperfectly reflect the meaning of constructs for different people. We have in some cases made inferences about which variables were directly associated, based on the assumed meaning of the variable. Some issues about the meaning of variables are discussed in the conclusion section.

The discussion section is farther from the data than are the conclusions and relies more strongly on the subjective interpretation of the authors. One major aspect of the discussion section is that we infer causal

relationships between variables. We talk about the possible meaning of variables and of the relationships among variables. An effort is made to give concrete examples of what these results mean for the practice of teaching. Limitations to the results are mentioned. Directions for theory development and future research are sometimes suggested. In writing this discussion, we have tried to rely on knowledge of other research, sound reasoning, and the common sense that comes from experience. Other peoples' interpretations may differ from ours.

Limitations to the Study

A number of additional methodological and substantive considerations bear on the interpretation on this study. These are outlined briefly in this section. The academic learning time model of classroom instruction relates three major sets of constructs: teaching processes, ongoing student learning behavior and student achievement. In transforming these constructs into observable and measurable variables, many choices were made. We attempted to choose teaching process variables, academic learning time variables and achievement variables which represented major elements of their respective constructs in ways that could be reliably measured and would have practical application. This process involved compromise. Teaching processes are extremely complex and hence, it is impossible to describe them completely with a relatively small number of measures. Similar statements can be made about academic learning time and student achievement. As a result, there are bound to be specification errors associated with our choices of measurable variables; hence the data are bound to distort the true relationships among the constructs.

Once measurable variables are chosen, it is extremely difficult to devise measurement procedures which are error free. In this study,

substantial measurement error occurs in most variables. Such errors are characteristic of the kinds of variables and the current measurement methodology used in the social sciences. Great care was taken to eliminate as much error from our measures as possible and, in several cases, this study has provided new and improved field methods for the study of classroom teaching and learning. In almost all cases, the size of measurement errors was estimated. The effect of such errors is to reduce the strength of relationships between measured variables. Thus the true relationships between variables are underestimated, depending upon the size of the measurement errors.

When considering the meaning of the Phase III-B findings for students who were not part of the study, we must look at the representativeness of the sample of teachers, classes, and students. All teachers who participated in the study were volunteers and all classes were selected from schools in the San Francisco Bay Area. The target students, selected in each class, represented a restricted range (approximately 20th to 60th percentile) on the overall distribution of student entering achievement. It is possible that somewhat different results would be found for samples of classes and students selected in a different way. In particular, somewhat different relationships may hold for students who are younger, older, or of different academic status from those in the study.

Most of the analyses of the data were conducted in a linear multiple regression framework. Hence, non-linear relations in the data are, in the main, undetected. The data were examined to look for some important interactions. Two and three way plots of data were constructed for this purpose. However, the search for non-linear relationships was by no means exhaustive.

Some types of interactions were ignored in the very design of the

study. In particular, we did not investigate the role of entering student characteristics, other than general academic status. It is unlikely that one type of instruction will be optimal for all students. We have not pursued such interactions in this study.

Conclusions

Academic Learning Time (ALT). Analyses of the relationship between ALT and student achievement were reported in Chapter 4. The strongest results were obtained when the components of ALT were included as separate variables. Conclusions are stated by component. Chapter 9 included analyses of the relationship between ALT variables and student attitude toward reading, mathematics, and school.

Two groups of conclusions about ALT are presented. The two groups differ in the straightforwardness of inference from the data and therefore in the degree of confidence with which they are presented.

1. Time allocated to instruction in a content area is positively associated with learning in that content area.

Across the 52 regressions analyzing the combined effects of the ALT variables (Tables 4-12 through 4-15), the amount of time allocated to instruction was significantly associated ($p \leq .10$) with residual student achievement in all four quadrants;² and significant associations were always positive (with one exception). The unique effects for allocated time appeared more consistently in the B-C period analyses than in the A-B

²The term quadrant is used to refer to one of the four cells of the design -- grade 2 reading, grade 2 math, grade 5 reading, or grade 5 math. The same sample of students was followed in reading and math at each grade. Note that several regressions were run in each quadrant, so that "significantly associated...in all four quadrants" does not mean significantly associated in all regressions in all four quadrants.

analyses. In the A-B analyses, significant positive effects for allocated time appeared in the grade 2 reading and grade 5 mathematics quadrants. However, inconsistent effects (one positive and one negative) appeared in grade 2 mathematics and no significant effects appeared in grade 5 reading. In the B-C analyses, significant, positive effects for allocated time appeared in all four quadrants.

2. Student engagement rate during allocated instructional time in reading or mathematics is positively associated with learning.

Across the regressions analyzing combined ALT effects, engagement rate was significantly associated with residual achievement in all four quadrants; and significant associations were always positive. The unique effects for engagement rate appeared more consistently in the A-B analyses than in the B-C analyses. Significant, positive effects for engagement rate appeared in all four quadrants in the A-B analyses. In the B-C analyses, significant, positive effects for engagement rate appeared in all quadrants except grade 2 mathematics, where no significant effects occurred.

3. The proportion of allocated reading or mathematics time spent by students on tasks where they have low error rates is positively associated with learning.

Across the regressions analyzing combined ALT effects, low error rate was significantly associated with residual achievement in all four quadrants; and significant effects were always positive. The unique effects for low error rate appeared more consistently in the A-B analyses than in the B-C analyses. Significant, positive effects for low error rate appeared in all four quadrants in the A-B analyses. In the B-C analyses, significant, positive effects for low error rate appeared in all quadrants except grade 5 reading, where no significant effects occurred.

4. The proportion of allocated reading or mathematics time spent by students on tasks where they have high error rates is negatively associated with learning.

High error rate was significantly associated with residual achievement in all four quadrants; and significant effects were always negative, across the regressions analyzing combined ALT effects. The unique effects for high error rate appeared more consistently in the B-C analyses than in the A-B analyses. In the A-B analyses, significant, negative effects for high error rate appeared in the grade 2 mathematics and grade 5 mathematics quadrants, while no significant effects appeared in grade 2 reading or grade 5 reading. In the B-C analyses, significant, negative effects for high error rate appeared in all four quadrants.

5. Greater quantities of ALT are not associated with declining student attitudes toward reading, mathematics, or school.

Table 9-5 indicates that the four ALT variables were not negatively associated with residual student attitude scores (recall that less high error rate indicates greater ALT). Time allocation and engagement rate showed no significant ($p \leq .05$) correlations with residual student attitude scores and no consistent pattern of nonsignificant correlations. Low error rate was positively associated with residual attitude for all twelve correlations computed, with two of these correlations reaching significance. High error rate was negatively associated with residual attitude for all six grade 2 correlations, one of which was significant. However, high error rate showed a mix of positive and negative associations with residual attitude at grade 5, where the two correlations that reached significance were both positive. Therefore, the findings for high error rate at grade 5 do not support the general trend, in that a lack of ALT shows some association with improving attitudes. Nevertheless, the general pattern did not

indicate that ALT is associated with declining student attitudes.

The following conclusions are relatively more arguable than the preceding conclusions. It is the belief of the authors of this report that the results of this study support the following conclusions. However, we wish to recognize that others will interpret the findings differently in some cases.

6. The association of the combined ALT variables with student learning is of considerable practical importance in terms of student learning in schools.

The practical importance of the ALT and learning association is, in part, indicated by the proportion of variance in residual achievement that is accounted for by the combined effects of the ALT variables. Out of 208 possible effects for the four ALT variables, 73 effects were significant ($p \leq .10$). Of these 73 effects, only one ran contrary to the general trends for these variables. This contrary effect occurred for allocated time in addition and subtraction speed tests at grade 2 during the A-B period. However, during the B-C period the same time category was positively associated with learning and accounted for more variance than its negative A-B period effect.

In addition to the consistency of the findings for ALT, the combined effects for ALT were found to account for an average of about 6% of the A-B period residual achievement variance and 11% of the B-C period variance. Furthermore, it is likely that these proportions of variance are considerably reduced by error of measurement, so that the "true" association of ALT and learning is much stronger.

In addition, examples of the predicted effects of changes in ALT variables such as allocated time, based on linear extrapolations from the data of this study, indicate that important changes in student learning

could be associated with intentional changes in ALT. Such examples are purely speculative, but they do show that the magnitude of effects found for ALT could be of considerable practical importance.

7. The association of low error rate with student learning is due to that part of low error rate that is an instructional treatment, rather than the part of low error rate that reflects student skills and abilities.

It is not possible to clearly determine whether low error rate was positively associated with residual achievement simply because it reflects student skills and abilities or because it represents certain aspects of instruction (an instructional treatment). However, there were some clues suggesting that the association of low error rate with residual achievement was probably due to that part of low error rate that is an instructional treatment. One clue was lack of correspondence between the size of effects for low error rate and the size of effects for pre achievement (which is a measure of entering student skills and abilities). This suggests that the size of effects for low error rate are not simply a function of the size of effects for student skills and abilities that remain unaccounted for by pre achievement.

A second clue is represented in Table 4-16, which indicates that where the proportion of time in low error rate and pre achievement scores are both low, both medium, or both high, agreement between these two measures did not exist in terms of predicted post achievement. This was particularly true where both the proportion of low error time and pre achievement scores were low. This may suggest that low error rate was associated with residual achievement because it measured something other than the skill and ability

domain assessed by pre achievement.

Finally, the fact that teaching process variables were related to low error rate suggests that error rate is influenced by teaching processes and is, therefore, an instructional treatment variable.

8. The effects of low error rate are greater for those students who are lower in pre achievement.

The possibility that low error rate has a greater effect among students who have lower pre achievement scores is indicated by Table 4-16. The difference in residual achievement between the low and high level of low error rate (representing the "effect" of low error rate) is generally greater within the low pre achievement rows (of Table 4-16) than within the medium or high pre achievement rows. In fact, in grade 5 reading and mathematics, low error rate appeared to have no consistent effect among students who were high in pre achievement. This may indicate that low error rate has a greater effect among low pre achievement students and almost no effect among high pre achievement students at grade 5. These conclusions, should be considered tentative.

Teaching process variables. In Chapters 5 and 6, we introduced teaching process variables reflecting the five functions of diagnosis, prescription, presentation, monitoring and feedback, as well as ratings of general classroom context. We looked at the relationship of these variables to student engagement rate (Chapter 5) and student error rate (Chapter 6). In Chapter 7, we considered how teaching process variables and ALT together relate to student achievement. From the analyses in Chapter 7, it became apparent that some of the BTES teaching process variables were related to student achievement in ways that do not coincide with ALT as now measured. In other words, both ALT and residual achievement as we measured them are

useful measures of student learning and they measure somewhat different aspects of learning. In the conclusions that follow, the most general effects are described as associations with "student learning." Specific outcomes are identified where appropriate. Nine major conclusions are listed below, in the order of the five functions.

1. Teacher awareness of the academic strengths and weaknesses of individual students (diagnosis) is positively associated with student learning.

Teacher diagnostic skill was measured by asking teachers to predict how students would do on representative items from the BTES test battery. Accuracy of teacher item prediction was significantly positively correlated with residual achievement in three of four correlations for total reading or total mathematics.³ The relationship between item prediction and achievement was mediated in part by student error rate. Item prediction was negatively correlated with high error rate in seven of eight correlations; two correlations reached statistical significance ($r = .45$ at the class level). The negative relationship between item prediction and high error rate was particularly strong in the A-B period. Item prediction was also associated with student engagement. Three of four correlations were positive; two of these were statistically significant.

2. Prescription of instruction which is appropriate to the needs and level of performance of individual students is positively associated with student learning. Appropriateness of prescription is related to the student error rate component of ALT.

³Unless otherwise noted, correlations with teaching process variables are labeled "statistically significant" when the correlations are .18 at the student level ($p \leq .05$ for a single correlation).

The appropriateness, or reasonableness, of the instructional program, relative to the needs of individual students, was rated by field workers. This rating appears to reflect the degree to which the teacher or the instructional system as a whole uses diagnosis on an ongoing basis to prescribe instructional treatments for students. Appropriateness was positively correlated with residual achievement in three of four correlations for total reading or total mathematics; one of the correlations reached statistical significance. Appropriateness had complex relationships with error rate (see Chapters 6 and 7). Appropriateness was consistently negatively associated with high error rate. Eight of eight correlations were negative; the relationship was stronger and reached statistical significance in grade 5 ($r = .45$ at the class level). Appropriateness was also significantly negatively associated with low error rate in grade 5. That is, there was proportionally more medium error rate in grade 5 classes rated higher on appropriateness. Interpretations for this relationship are offered in Chapters 6, 7, and later on in this chapter.

3. More substantive interaction between the student and an instructor is associated with higher percentages of student engagement.

Substantive interaction is a general term for direct interaction between the student and an instructor about academic content. It includes both one-to-one interaction and interaction between the instructor and a group of students in which the target student is either an active or a passive participant. The observed proportion of a student's instructional time during which s/he was involved in substantive interaction was strongly correlated with student engagement (about $r = .45$, at the student level, in all four quadrants).

Substantive interaction is composed of behaviors representing the

interactive functions of presentation, monitoring, and feedback. Both academic monitoring (questions) and academic feedback were significantly positively correlated with engagement in all four quadrants. Presentation variables, explanation of concepts and structuring/directing, were significantly positively correlated with engagement in mathematics.

Analysis of relationships by instructional setting indicated that most interactive instruction took place in a group setting, with a relatively small part coming from one-to-one-tutorial during seatwork. Analyses of setting indicated that engagement rates were higher when students spent more time in a group setting. However, amount of substantive interaction was a more potent predictor than amount of groupwork. When group time was more nearly filled with substantive interaction, as opposed to organizational tasks or waiting for others, then engagement rates were higher during groupwork and during seatwork. When students received more contact with an instructor during seatwork, then engagement rates were higher in seatwork. Engagement rates were especially low when students spent two-thirds or more of their time in seatwork and received little attention. The description of classes in Chapter 8 suggested that use of aides, parent volunteers, cross-age and peer tutors was associated with increased amounts of interactive instruction for students.

4. Structuring of the lesson and/or giving directions on task procedures is/are positively associated with low error rate.

The observation category labeled structure/direct was a combination of talking about the structure of the lesson and giving directions about what the student should do. As with all observation variables, we calculated the

percentage of the student's instructional time during which an instructor was providing structure/direction. Higher percentages of structure/direct were associated with higher percentages of low error rate in all four quadrants; one of four correlations reached statistical significance. This result is discussed in Chapter 6.

5. Explanation specifically in response to student need is negatively associated with low error rate and positively associated with high error rate.

The percentage of instructional time during which students received explanation specifically in response to need was assessed by direct observation. This variable was negatively correlated with low error rate in all four quadrants. It was positive correlated with high error rate in all four quadrants. Four of the eight correlations, two positive and two negative, reached statistical significance. This result is discussed in Chapter 6.

6. Academic feedback is positively associated with student learning.

Academic feedback was defined as information to the student about whether his or her answers were right or wrong, including giving the right answer. Many different specific behaviors were conceptualized as fulfilling this function, including student answers to teacher questions in a group lesson and oral reading. The percentage of instructional time during which the student was receiving feedback was positively correlated with residual achievement in all four quadrants; two of the four correlations were statistically significant. Part of this relationship is mediated by student engagement. As reported in point 3, frequency of academic feedback is positively associated with student engagement. But the association between academic feedback and engagement is not strong enough to explain the whole association between academic feedback and achievement. Correlations between

academic feedback and error rate were all insignificant. Possible explanations are discussed in Chapter 7.

7. More frequent negative feedback to students about their behavior is associated with lower student engagement rates.

Task engagement feedback had a consistent, negative association with engagement. Correlations were statistically significant in all four quadrants. Task engagement feedback turned out to be primarily negative feedback, that is, it occurred at moments when students were off-task (statements like "Johnny get back to work.") Task engagement feedback was given more often to students who had generally lower rates of engagement. It was given more often on days when the student was more frequently off-task.

8. Teacher emphasis on academic goals is positively associated with academic learning. Teacher orientation toward affect to the exclusion of academic instruction is negatively associated with academic learning.

Several field worker ratings were averaged to get an estimate of the degree to which teachers attended to and valued positive affect. This variable, orientation to affect, was adjusted by partialling out an estimate of observer halo. This adjusted rating reflects the degree to which the teacher focuses entirely on affect, at one extreme, or entirely on academics at the other extreme. Primary orientation toward affect was associated with lower time allocation to reading and math (all four correlations negative; two significant), lower student engagement rates (all four correlations negative; three significant) and lower student achievement (all four correlations negative and marginally significant). Orientation toward affect tended to be positively correlated with high error rate but had mixed

correlations with low error rate. Another rating, perceived academic competence, which included academic emphasis, tended to be positively correlated with achievement (all four correlations positive; one significant). Perceived academic competence was also associated with higher student engagement in grade 5. In addition to these correlations, the descriptions of classes in Chapter 8 provided concrete examples. Three grade 2 classes with above average achievement were described in some detail and in all three the teacher clearly emphasized the importance of school learning. In contrast, one class with below average achievement clearly placed primary emphasis on other outcomes such as social and artistic development. This is not to say that the high-achieving classes were "cold" or ignored student feelings; but they did place primary emphasis on academics.

9. A learning environment characterized by student responsibility for academic work and by cooperation on academic tasks is associated with greater learning.

Following each observation, the field worker rated two aspects of the classroom environment. One was student responsibility, the degree to which students took responsibility for their classwork, belongings, and classroom. The other was cooperation, the degree to which students helped each other, shared materials, and worked together. These two ratings were highly intercorrelated, so they were averaged to form one variable for analysis. This variable, learning environment, was significantly positively correlated with achievement in grade 2. Both components, cooperation and student responsibility, contributed to this relationship. Descriptions of specific classes indicated that there was an interaction with academic focus. In classes where students cooperated on and took responsibility for academic

work, as opposed to classroom routines like taking attendance, achievement was greater. In grade 5, student responsibility was strongly associated with academic "press" from the teacher (perceived academic competence). Student responsibility was significantly positively correlated with student engagement and achievement in reading. Cooperation was not significantly correlated with achievement in grade 5.

Discussion (The Model Revisited)

This section presents discussion and comments on the constructs in the ALT model in light of the findings of the study. This discussion is "farther" from the data and includes a greater amount of speculation and opinion than the previous section in this chapter.

Academic Learning Time (ALT revisited). The Academic Learning Time model is the product of a long evolution of ideas. It started with the notion that instructional time ought to be a key variable underlying the relationship of schooling processes and learning outcomes. Next, instructional time was divided into two components: allocated time and engagement during the allocated time. Sometime later, the notion appeared that some students might learn more or less as a function of their ability to perform the instructional tasks that they were given. This led to the categories of low, medium, and high error rate. Initially, we expected to find that medium error rates would produce greater learning. However, others pointed out that low error rate, in the form of practice and review, can be a useful and perhaps necessary part of the learning process. This persuaded us to look for a balance between low and medium error rates as the most effective instructional process. Analyses of BTES data, however, led to the conclusion that relatively low error rate instruction is beneficial, while relatively more medium and high error rate is detrimental.

This evolution of the ALT model is continuing. As a result of conducting the BTES study, watching and thinking about the experiences of students and teachers, pondering data that have been regressed and charted and plotted, our current ideas on ALT take a slightly different shape than the more constrained listing of conclusions based on data analyses (presented previously in this chapter). What follows is an attempt to express some of what we currently believe about "ALT." The reader should feel free to disagree with any or all of it.

The most important feature of ALT is that it attempts to represent ongoing student behaviors in the classroom that actually manifest learning itself. This is a conceptual strategy that is far more powerful than the traditional procedure wherein learning is inferred to have occurred sometime between two testing occasions, if the second testing score is higher than the first. By trying to think in terms of ongoing student behaviors that are, themselves, the learning process, then we are in a much better position to evaluate teaching processes.

The analyses relating ALT to post achievement residuals (a traditional criterion of learning) have been presented as a means of judging the ability of the student behaviors used in this study to represent ongoing learning. This should not be taken to mean that post achievement residuals are a more accurate representation of learning than ALT (or that they are not). Rather, achievement residuals, as a result of the considerable body of existing research that currently exists, are a highly refined and generally accepted criterion of learning. It is not suggested here that measures of ALT should ever replace achievement test scores. Rather, it is suggested that ALT should supplement achievement testing.

Our current view of ALT (for "ALT" one could read "the learning

student") is that it consists of active involvement in an activity that is directly relevant to some content area, where this involvement is generally successful. "Activity" refers to a learning opportunity that may be one of several options available to the student. The availability of activities, in terms of number, duration, response requirements, and content relevance, determines the potential for learning by a given student.

"Relevance to some content area" is a curriculum issue. Any given activity will be relevant to multiple learning outcomes. It is not always clear which activities are most relevant to a particular outcome. In any case, a fully refined model of ALT would attempt to specify several different content areas to which a given task is relevant and the degree to which it is relevant to each content area. "Active involvement" refers to the extent to which the student actually avails him/herself of a given learning opportunity. In its simplest form (which was used for this study), active involvement means that the student attempts to perform the activity. A more refined conceptualization of active involvement would consider a range of affective/motivational variables, such as the levels of concentration, enthusiasm, and anxiety with which the student attempts to perform the activity.

Note that active involvement and the availability of an activity probably interact. If the same activity is "available" for too long a period of time, then students may become saturated or bored, so that active involvement decreases. Therefore, it is important that the availability of an activity not be increased to such an extreme that active involvement is considerably reduced.

"Generally successful" refers to the error rate components of ALT. (Note that it would probably be better to say "success rate" instead of

error rate. Perhaps future work involving ALT will use this more positive term.) We have found that students learn more when they spend relatively more time working at a low error rate and relatively less time at a high error rate. This may be thought of as representing the student's comprehension of the activity being performed. That is, error rate refers to the cognitive (as contrasted with affective) variables involved in student learning processes. Presumably, learning will occur if the student's comprehension of the activity at hand is generally increasing. This increase in comprehension is learning. It appears that comprehension increases more rapidly when it is generally already at a high level, than it does starting at a very low level (relative to a particular task). Note that this conclusion should be represented in terms relative to what we found in typical classrooms at the second and fifth grades.

All this leads to the following view of what the "learning student" actually looks like. First, he/she has some activity or task that is designed to result in increased knowledge or skills. This may be writing answers to a series of problems or questions, reading a book silently or aloud, listening to a teacher's explanation, watching a demonstration, etc. The "design" of the activity is important, in that some are more or less effective than others for facilitating certain learning outcomes. The variety and sequencing of activities are probably important design factors in their relationship with learning outcomes. However, the present study does not say much about the specifics of activity design.

The learning student is very attentive. He/she is actively involved in the activity at hand, probably with some enthusiasm. The learning student is busy performing the academic part of the task, rather than sharpening his/her pencil, looking for a book, or waiting in line to ask the teacher

a question. He/she is not socializing or daydreaming. Nevertheless, the student is enjoying the activity. Paying attention for relatively long periods of time does not upset the student. Furthermore, his/her success on the task makes this active involvement more enjoyable.

The learning student spends a lot of time practicing and reviewing skills. He/she undertakes an activity related to a new skill only after thoroughly learning skills prerequisite to the new skill, so that he/she virtually never encounters an activity that is really entirely "new." At a younger age and/or an earlier stage of education, the learning student spends relatively greater periods of time practicing and reviewing. As he/she advances academically, it becomes possible to reduce the proportion of time spent practicing and reviewing. It appears that at some point (roughly at the level of an advanced grade 5 student), the learning student is able to spend more time attempting newer skills than on practicing previously learned skills. There is always some need for consolidation of acquired skills (practice), but as the student advances, he/she actually "learns how to learn," so that it becomes easier to acquire newer skills without as much consolidation of prerequisite skills. In addition, attempts to learn the newer skills probably serve to consolidate the previously acquired skills, insofar as the more advanced student is applying these acquired skills to the newer skill.

The need to spend relatively more time practicing and reviewing skills is probably partially a function of the fact that students are not constantly tutored. That is, students spend the majority of their time working independently or with only intermittent teacher contact. Students, particularly less advanced students, have difficulty acquiring new skills on their own, so they need to practice previously acquired skills when

working alone. Therefore, as students become more advanced, they should not only be able to spend less time practicing and reviewing, but could spend more time working independently. This would explain the fact that some highly capable students learn quite well on their own, without instruction by a "teacher."

It should also be noted that practice and review appears to enhance retention over a period of no instruction, such as a summer vacation (see Chapter 10). It is possible that the less advanced student is slower to internalize acquired skills and store them in memory (perhaps as a function of having fewer related pieces of knowledge with which to form associative webs). Thus, these students must practice and review in order to retain after instruction ends.

Of major importance is the conclusion that this "learning student" is not necessarily an unhappy student. The learning student does not learn to dislike learning. Lots of hard work may sound undesirable to most people. However, we do not find any evidence that students are less satisfied when the sheer quantity of work (allocated time) is relatively great. Furthermore, we do not find that students who pay more attention (work intensively) acquire a distaste for learning. In fact, one could assume that when attention is the result of interest and enthusiasm, rather than coercion, that it will represent a more positive attitude toward learning. Recall (from Chapter 5) that student engagement does not appear, generally, to be produced by the coercive demands of the teacher (negative task engagement feedback), as in "Get back to work."

It is interesting to note that the low error rate component of learning (success) is associated with more positive student attitudes. Successful students probably enjoy learning more because of their success. Failure, in the form of high error rate, appears to result in a more negative

attitude among younger students. However, this may not be true as students acquire more academic experience and become more accustomed to school. It is possible that more advanced students are not readily discouraged by periodic failure. This may relate to the more advanced student's ability to learn with more challenging tasks for which he/she experiences less success. Perhaps the ability to learn from initial failure reduces the aversiveness of failure, or possibly the more positive attitude toward failure experiences enables the advanced student to learn from them. More likely, both relationships have some validity.

What next? Perhaps the most exciting finding of this study is that learning processes that are, at least to some extent, under the control of the teacher account for very important proportions of achievement outcomes. Massive doses of allocated time, provided in such a way that engagement levels do not decline seriously, might be a highly effective tool for remediating failing students. Certainly, mastery learning advocates have been saying this for years.

Increasing instruction geared to lower levels of student error rate may be another effective instructional strategy readily available to most teachers. This strategy is particularly promising at earlier grades and for less advanced students. Again, note that there have been advocates of this approach in previous years (programmed learning and mastery learning). It is also possible, however, that most teachers do not recognize the extent to which less advanced students need practice and review (see discussion of teaching processes, below). If adults learn well with relatively little practice and review, then they may expect younger students to do as well. The relationship of low error rate to learning, then, may be particularly valuable information to most teachers at the early

elementary levels. A related strategy is the use of high error rate as a diagnostic flag, signalling serious instructional problems and the possible need for a simpler activity.

None of these applications of the ALT findings should be disseminated on a wide scale without considerable further investigation of them. One approach would be to conduct experiments in an attempt to more clearly determine causal relationships. Another approach would be to develop practical applications of these findings and to "try them out" in a few classrooms. It may also be useful to simply measure ALT variables for classroom teachers (or have them measure ALT themselves) as a feedback device. Perhaps a few able teachers would be the most qualified specialists for developing applications of the ALT findings. In any case, a variety of intentional manipulations of ALT variables (experimental, interventive, or teacher initiated) should be conducted in classrooms.

Having observed and contemplated, we should now tinker around a bit.

Diagnosis. The data support the conclusion that diagnosis is an important part of effective teaching. Students learn more when teachers know more about what their individual students can and cannot do.

We were primarily interested in cognitive achievement. To foster cognitive achievement, it is important for the teacher to know the cognitive skills and level of performance of individual students. Our measure of this teacher skill was to ask teachers to predict how their students would do on representative items from the BTES test battery. Students learned more when their teachers were more accurate in predicting performance. Teachers were more accurate in predicting cognitive performance when they knew more about the subject matter area and when they attended to differences between students. Presumably, knowledge of subject matter (reading and

mathematics pedagogy) allows the teacher to label critical aspects of performance, to know what makes one test item harder than another, to know whether or not students have had instruction on problems of that kind, to know what it is that the student doesn't understand. Abstract knowledge is not enough, however. The teacher needs to see differences between students, to see that Johnny has understood the concept being taught but Mary hasn't. Teachers are aware of many kinds of differences between students -- differences in cognitive performance, differences in personality, differences in work habits, etc. The fact that perception of differences was not strongly related to achievement suggests that it is important what kind of differences the teacher notices. It may be that it is particularly important to attend to differences in cognitive performance. When knowledge of subject matter and awareness of student differences are combined to give the teacher an understanding of what each student can and cannot do, then a foundation for instructional planning has been established.

In the BTES study, we did not study systematically how teachers who were good diagnosticians got their information. Interviews with teachers early in the year indicated that they used a variety of formal and informal assessment techniques. Some teachers gave commercially published tests; some had devised their own tests; some used special classroom activities designed to give the teacher information about what students could do. Teachers also had different ways of keeping in touch with student performance during the year. Presumably, testing can provide an opportunity for monitoring progress and updating one's diagnosis. Perhaps more importantly, the student's responses to daily tasks provide an opportunity for informal evaluation. The positive outlier classes described in Chapter 8 were characterized by monitoring of daily performance, feedback

to students, and positive rewards for accomplishment. While carrying out these instructional functions, the teachers presumably learned a lot about the students.

Prescription. Prescription refers to the process of deciding what students should work on in the classroom. This is a complex area and one with mixed results in the data.

The major positive factor in this area was the "appropriateness" of the instructional program for the needs of the students. This variable represents the integration of the two planning functions -- diagnosis and prescription. It assesses the extent to which teachers use their knowledge of individual students to prescribe apparently reasonable instructional programs. Appropriateness was related to error rate and to achievement.

Our measure of appropriateness was a rating made by the field workers based on interviews and extensive observation. It is difficult to pinpoint exactly what was considered appropriate. The instructions on the rating form and supplementary comments by field workers provide some suggestions. Field workers were asked to consider the pacing of instruction, whether faster students could move ahead while slower students received extra help. They were asked to consider student error rate. A hypothetical example of appropriateness was that the teacher might notice a student's interest in mathematical puzzles and bring in some additional materials s/he might be interested in. A hypothetical example of inappropriate instruction was that the teacher had all students in the same reading book regardless of clear differences in reading skill. Field workers were asked to give the reasons for some of their ratings. The most salient dimension in grade 2 classes appeared to be flexibility of grouping. When giving instances of appropriateness, field workers often commented that the teacher would

"regroup students according to needs" or that a student who was doing particularly well or particularly poorly was moved to another group. For grade 5 classes, this same dimension appeared, but field workers also seemed to attend to the overall organizational structure of the class. In grade 2, most teachers have some form of small group instruction at least part of the time. In grade 5 there is more diversity. Some classes use whole class instruction; others have small group instruction; others have individualized programs where students work independently. In grade 5, individualized programs tended to get high ratings of appropriateness, since students were usually assigned tasks based on their individual test performance and could move ahead at their own rate.⁴ One grade 5 teacher was given a moderately high appropriateness rating for forming reading groups and then adapting the curriculum through the year to the level of performance, interest, and pace of each group.

One thing that should be clear from these examples is that the definition of "appropriate" is fairly broad and general. It assesses not whether each student was given the best instruction for his or her needs (something that would be impossible to determine) but instead whether the program appears to be reasonable for the different students in the class.

This description of appropriateness as reasonableness is consistent with the correlations we obtained between appropriateness and student error rate. Classes rated more appropriate had a smaller percentage of high error rate tasks. In other words, it is generally unreasonable and

⁴ Individualized programs such as this were also considered highly goal directed, since they were organized around skill areas and used testing to monitor and regulate student progress. This probably explains the high correlation between appropriateness and goal orientation in grade 5. Individualized programs also place more responsibility on the student, as shown in the correlations between learning environment and appropriateness or goal orientation.

inappropriate to give students tasks that they cannot do. At least, if tasks are too hard either the curriculum or the grouping should change.

At the boundary between low and medium error rate, appropriateness is less clear. In grade 5, classes rated high on appropriateness spent relatively more time at a medium error rate than did classes rated less appropriate. It may be that this happened because of the nature of the individualized programs that were rated high on appropriateness. Students work on their own and thus have to figure things out on their own. They are given tasks that they have not yet mastered but which they do have the necessary skills to figure out. As soon as they can perform the task correctly, they move on to the next stage. In other words, there is very little review and practice at the low error rate level (successful practice). Group instruction may tend to involve more easy work because this makes it easier to handle the group and/or because faster students overlearn while slower students are reaching mastery. It should also be noted that in grade 5 both appropriateness and a higher proportion of medium error rate were associated with perceived academic competence. Some teachers stressed academics, provided competent instruction taking into account individual student performance levels, and expected a lot of students. Descriptions of one such class, a high-achieving class, commented on the challenging academic curriculum developed by the teacher and on the teacher's demand for responsible, work-oriented behavior from students. In sum, the proportion of low error rate seems to be a function of the type of classroom organization and/or of what the teacher considers to be appropriate.

The above discussion suggests that it may be useful to think of error rate itself as a prescription variable. Diagnosis followed by reasonable prescription will guide the teacher away from giving tasks that are too

hard. Only a conscious and perhaps counterintuitive effort to provide low error rate practice will produce more low error rate. Especially in grade 5 it seems imperative to work with teachers in an experimental sense to determine the consequences of such a program.

The above discussion also suggests the need to study classroom task structure, grouping, and curriculum. A decision to include more low error rate tasks (successful practice) may be implemented by changing the pacing built into the curriculum (or changing to a different curriculum). It may require changing the number or composition of instructional groups so that pacing can be better adapted to individuals. Decisions about grouping and curriculum are the substance of prescription. They have not been explicitly or thoroughly studied in BTES, although there are indications of the mediating role of such factors. This is an important area for further study.

Some comments are in order about other aspects of prescription. One factor that we did not examine in the BTES study is the issue of who decides what should be done. The five function model assumes to some degree that the teacher is a decision maker. The logical progression of diagnosis leading to prescription leading to classroom events implies that the teacher is actively using knowledge of students to help decide what should be done. This may suggest a traditional classroom model where the teacher determines who does what when, but this is not required by the model. There is room for students to participate in choosing activities, setting goals, and determining timelines. The teacher's role as guide and participant is more complex and subtle in such an environment than in the traditional classroom, but still it is possible and presumably desirable for the teacher to get to know the students and to present choices or introduce ideas based on what is known about student interests and needs. One can argue that instruction

is more likely to be appropriate to the individual student when the student plays an active role in decision making. We did not explore this issue in the study. The classes in the study tended to be relatively teacher-centered. They ranged from classes where the teacher made all decisions (note that both whole class instruction and individualized programs can fall in this category) to classes where students and teacher negotiated about what was to be done (within a relatively narrow set of alternatives). The appropriateness variable estimates to some extent whether student needs were taken into account in planning the program. It is relatively independent of the issue of whether students actually helped make decisions. One can imagine both positive and negative aspects to student choice, so this is an area for further study.

Presentation. Presentation skills appear to be useful for increasing student engagement in mathematics. Teachers tend to explain concepts more often in mathematics than in reading. The term "explain" is used broadly here. Demonstrating the steps involved in an addition problem would be considered explanation. Students pay attention more in mathematics when they receive more frequent planned presentation of concepts in a group setting.

They also pay attention more when the teacher spends time discussing the goals or structure of the lesson and/or giving directions about what the students are to do. Perhaps because of the tendency to give more seatwork in math than in reading and because of the variety of problems to work, it is important that students know the context of the lesson and what they are to do, so that they become more involved in the task.

In both reading and math, students tend to make fewer errors on daily tasks when teachers spend more time structuring the lesson and giving

directions. It may be that students make some errors because they don't understand the format of the page they're working on or don't recognize the skill they're supposed to apply. Student workbooks are notorious for their frequently changing formats and confusing directions. It seems important that students understand what they are supposed to do so that they can respond correctly.

Descriptions of particularly successful classes often mentioned that the teacher had a regular routine of beginning each lesson with a presentation in a group setting. The teacher would tell the students what they were going to work on, make sure all students understood the assignment, go over examples where appropriate. The model math lesson used by Good in his experimental study (1978) began in a similar way. This seems like an efficient and productive procedure.

One presentation variable was consistently negatively associated with low error rate and positively associated with high error rate. This variable was explanation-need. In effect, students who made more errors and did not understand classroom assignments received more explanation specifically in response to need. Students who needed help, got help. While this seems reasonable, one should be wary of overreliance on this teaching behavior. Explanation should increase understanding and reduce the overall error rate. In our classes, this tended not to happen, hence the negative association. The danger is that explanation-need is too little, too late. Frequent use of explanation-need might best be interpreted as a symptom that the level or pacing of instruction are inappropriate for the student. Major changes may be in order.

Monitoring. As measured in the BTES observation system, the major form of monitoring was teacher questioning in a group setting. Teacher

questions account for about one-third of the interactive, substantive instruction that takes place. Students pay attention more when they are more often involved in substantive interaction, and teacher questions are an important part of that process. They involve the students in the interaction as well as giving the teacher information about what the students understand.

We did not examine the type of questions or pattern of questioning. Questions could be addressed to the group, to the target student, or to another student. The teacher could call on volunteers or go around the room in ordered turns. As long as the intention was for the target student (usually for all students) to listen to the question and think of an answer, a question was coded. Presumably some types or patterns of questioning may be more effective than others for getting all students to respond in appropriate ways.

The term "monitoring" can also be used to refer to the teacher behavior of circulating around the room during seatwork checking on how students are doing. We did not code general surveillance of the classroom. We did find that a teacher rarely stops to observe a student's work without making some comment, in which case we coded the comment. We also found that when a student does receive more attention from an instructor during seatwork (feedback or explanation) s/he pays attention more. Thus, monitoring seatwork by going around the room giving help or feedback as frequently as possible is a good idea. Accounts of high-achieving classes suggest that good teachers do this not only to keep kids on task but also to find out as much as they can about how students are doing, so that they can plan further instruction.

Feedback. Academic feedback to students should be provided as often as

possible. When students receive more frequent feedback about whether their responses are right or wrong, or hear the right answer, they pay attention more and learn more.

Academic feedback as defined in the BTES observation system includes many different behaviors. We don't know at this point what types of feedback may be more valuable than others. We can at least suggest some of the possibilities.

As defined in BTES, feedback is the major component in group interaction. Much classroom interaction follows a question-and-answer or recitation format: the teacher asks a question; a student answers the question. Presumably, when the teacher asks a question all students are supposed to think of an answer. When some student gives an answer orally, each student gets feedback on his or her own internal answer. When a student was giving an answer to a question, this was coded as feedback to students listening to the answer. This kind of feedback probably accounts for a large part of the positive relationship between feedback and student engagement. That is, it may not be feedback per se, but feedback as part of substantive interactive instruction that increases engagement. When students spend more time in substantive interaction, including some explanation, teacher questions, student answers, student questions, teacher and student comments, then student engagement rates are higher. Substantive interaction serves as a motivator to get students involved in classroom activities.

The above discussion suggests that feedback is one of several equal parts of academic interaction. Of the interactive behaviors, however, academic feedback was the most strongly and consistently positively

correlated with achievement. This suggests that there is something especially important about feedback and that many different forms of feedback in addition to question-and-answer recitation may contribute to learning.

Feedback was the one behavior which we coded when it came from curriculum materials rather than a human instructor. Programmed texts are organized to provide immediate feedback. Students might also check their answers in the back of the book or with an answer key. Feedback from the curriculum was not frequent in our classes, but it may be an important way to increase the amount of feedback students receive.

Students and other "untrained" instructors can provide feedback. To plan instruction, to choose learning tasks, or to explain concepts, requires some skill as an instructor and probably some training. Anyone who knows the answer to a problem can tell a student whether an answer is right or wrong or give the answer to the student. If an important part of learning is responding and receiving feedback, then all classroom personnel should be used wherever possible to provide feedback.

Some classes that we observed had a regular routine on group assignments of meeting to check answers as a group. One fifth grade math class had regular homework assignments and spent the first part of each day going over the homework. Presumably, students are more likely to complete tasks when they know they will be held accountable. A routine of checking papers gives everyone regular feedback. Homework and checking homework were among the most consistently implemented parts of Good's (1978) effective model math program.

Oral reading circle was a situation that we coded as involving high levels of feedback. Much like the recitation sequence, oral reading was

considered feedback to a student who was reading along silently. The teacher might also correct errors, thus providing feedback to all students.

In addition to academic feedback, we also looked at task engagement feedback, feedback to the student about whether his or her behavior was acceptable or unacceptable. Most of the task engagement feedback we observed turned out to be negative feedback, such as reminders to students to get back to work when they were off-task. We found no evidence that frequent use of such statement had any positive effects. It seem that when students are off-task a lot this elicits teacher reproaches. It may be that some well-timed and well-phrased reminders are useful, but when task engagement feedback becomes frequent this is a sign that some structural changes are needed. There is an important lesson here for teachers who try to utilize BTES findings and work to increase student engagement: yelling at kids more often is not the answer. (Instead, one might check to see that tasks are not too hard for the student (task engagement feedback was positively correlated with high error rate), increase the clarity and emphasis with which expectations are stated and the consistency with which students are held accountable, increase the amount of substantive interactive instruction, etc.)

A final comment comes from the Chapter 8 descriptions rather than the formal observation data. All three high-achieving classes described in that chapter had some type of positive reward system -- bionic handshakes, ice-cream cones, teacher praise, tokens saved up to get a prize. The concept seemed to be that good work deserved a reward. Such rewards were usually not frequent; but students had some sense, formal or informal, of what they had to do to get them. Such a system is on a larger time scale and is supplementary to our observed behaviors of academic feedback and task engagement feedback. Students need to know regularly and immediately whether their

answers are right or wrong, or what the right answer is. Teachers may find it useful to encourage desired behavior by comments such as "I like the way you're working quietly" or "I see you've finished the first page." There is also a place for reward systems which acknowledge major learning events; they give the student recognition for achievement, for working (perhaps working hard), and for succeeding.

Context. Schooling has many different purposes. One purpose is cognitive learning. Others might be development of independent work habits, learning social interaction skills, feeling good about oneself, enjoying work, appreciation of fine arts, keeping kids off the street, etc. Most teachers value and work toward a number of different outcomes. Because the BTES study focused primarily on cognitive outcomes, we cannot fully evaluate classroom instruction. Our data do point out, though, that choices must be made and that teachers should be aware of the choices they make.

Two of the general variables in the BTES study had to do with the focus or orientation of the teacher. One was academic orientation, the extent to which the teacher emphasized, valued, and worked toward cognitive achievement. The other was orientation toward affect, the extent to which the teacher was aware of, acknowledged, and valued student feelings.

Examining these variables in relationship to student engagement and student achievement reinforced an old maxim: "First things first." If the goal is to maximize student learning of basic skills, then it is important that the teacher value this goal and be committed to achieving it. The teacher must be willing to allocate classroom time to academic instruction and must communicate to the students the belief that academic learning is important. The teacher must be willing to make a personal effort to reach this goal.

The class descriptions in Chapter 8 illustrate this point. The high-achieving classes all had an emphasis on academics. In contrast, teacher F placed greater emphasis on nonacademic outcomes and devoted attention to music, personal development, and good feeling. There was little time spent on academics and achievement was low.

It seems reasonable to look for a middle ground. Many teachers give first priority to academic instruction but are also aware of student feelings and value human development. The high-achieving grade 2 classes show this to some extent. There are other good examples in both grades of teachers who make a sincere effort to provide competent academic instruction and also take into account student interests and human feelings. Often these classes are in the middle range of ALT and/or achievement. The conflicting demands of different goals may tend to moderate achievement gain. For many people this is an acceptable compromise.

In our sample there were grade level differences in the range of classroom environments. In grade 2, almost all classes had a relatively high degree of concern for affect. Some classes, like F, were oriented primarily toward affect. Even the most academically inclined classes had at least a moderately warm classroom environment. In grade 5, the sample as a whole was relatively more concerned with academics. No classes were as affectively oriented as class F. Some classes were very strict and exclusively work oriented. In these classes, field workers tended to describe the atmosphere as "tense," or "uncomfortable." These classes tended to score well on achievement tests. (Perhaps they were most thoroughly disciplined to try hard on boring tasks.) Some people might judge the cost too great.

Another context variable, "learning environment," needs to be

considered. This variable was a composite of two ratings -- one on "cooperation," the other on "student responsibility." In grade 2, classes higher on learning environment tended to have higher achievement. Both components, cooperation and student responsibility, contributed to this effect. Students learned more when there was a spirit of cooperation and where students took responsibility for classroom activities. An examination of individual classes indicated that this effect interacted with cognitive focus. Where students worked together to reach academic goals and where they took responsibility for achieving them, then achievement was higher. Cooperation and student responsibility in nonacademic pursuits did not have this effect. The high-achieving classes in Chapter 8 give examples of cooperation and student responsibility in academic pursuits.

In grade 5, the learning environment variable was less consistently and more complexly related to achievement. It was the student responsibility component that carried the relationship. It appeared that teachers with a strong academic orientation tended to expect and/or "demand" conscientious performance on the part of students. Students in these classes were very responsible. They knew what was expected and they knew what would happen if they didn't do their work (for instance, having to stay after school). The spirit of cooperation was missing in these classes. Cooperation was not related to achievement in grade 5. Perhaps those classes which emphasized cooperation placed relatively greater emphasis on affect. There is no reason to believe that fifth graders would not profit from an environment that stressed academic learning and created a sense of cooperation and student responsibility in pursuit of that goal.

One can construct from these results an image of an ideal class: a clear focus on cognitive learning; the students expect to work and are held

responsible for doing so; the teacher cares about the students and wants to help them learn; teacher and students interact comfortably and frequently on work activities. In other words, a class where the teacher emphasizes the belief that the purpose of school is learning and fosters an environment where everyone, teacher and students, works together to reach that goal.

A Last Word

Student learning is an elusive phenomenon. Teachers work hard to encourage it. Students work hard to achieve it. Researchers work hard to measure it and to understand how it comes about. We believe that many students learn a lot in school, and that the BTES project helps explain that process so that some students can learn more. We believe that there is a real advantage to being able to measure student learning as it takes place in the classroom rather than through formal achievement testing. Academic Learning Time (ALT) represents one attempt to measure ongoing student learning through observable classroom behavior. ALT has been validated as a measure of learning through its correlations with achievement. We hope that other people, especially teachers, will experiment with observing students to get immediate feedback about the learning process. ALT as we measured it suggests what to look at; we hope the construct will be used, adapted, and expanded by others as well as ourselves. The teaching functions and context variables that we examined suggest important aspects of instruction that relate to student learning. Here, too, we hope the model will prove useful and will be refined as people continue to analyze instruction. Most of these ideas are not new. None of these ideas are consistently put into practice. We hope this model will help some people to steer a course toward more productive education for students.

References

- Bloom, B.S. Stability and change in human characteristics. New York: Wiley, 1964.
- Carroll, J.B. A model for school learning. Teachers College Record, 1963, 64, 723-33.
- Dunkin, M.J., & Biddle, B.J. The study of teaching. New York: Holt, Rinehart and Winston, 1974.
- Gage, N.L. (Ed.). Handbook of research on teaching. Chicago: Rand-McNally, 1963.
- O'Connor, E.F., Jr. Extending classical test theory to the measurement of change. Review of Educational Research, 1972, 42, 73-97.
- Rowley, G.L. The reliability of observational measures. American Educational Research Journal, 1976, 13 (1), 51-60.
- Thorndike, R.L. (Ed.). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.
- Travers, R.M.W. (Ed.). Second handbook of research on teaching. Chicago: Rand-McNally, 1973.
- Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Reference Notes

1. Bloom, B.S. Time and learning. Address presented at the 81st annual convention of the American Psychological Association, Montreal, 1973.
2. Burstein, L. Assessing differences between grouped and ungrouped coefficients. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.
3. Cronbach, L.J. Research in classrooms and schools: formulation of questions, design and analysis. Stanford, California: Stanford Evaluation Consortium, School of Education, Stanford University, 1976.
4. Evertson, C. Personal communication, March 25, 1978.
5. Good, T. The Missouri Mathematics Effectiveness Project: a program of naturalistic and experimental research. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.
6. McDonald, F.J., & Elias, P.J. Beginning Teacher Evaluation Study: Phase II Final Report, Volume I. Princeton, New Jersey: Educational Testing Service, November 1975.
7. McDonald, F.J., & Elias, P.J. Phase II technical summary report: Beginning Teacher Evaluation Study. Princeton, New Jersey: Educational Testing Service, 1976.
8. Ogden, J., Brophy, J., & Evertson, C. An experimental investigation of organization and management techniques in first grade reading groups. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977.
9. Wiley, D.E. Another hour, another day: quantity of schooling, a potent path for policy. Chicago, Illinois: Study of Educative Processes, No. 3, University of Chicago, 1973.
10. Wright, C.E., & Kim, Y. A stability study of the student achievement tests for the Beginning Teacher Evaluation Study. Menlo Park, California: Educational Evaluation and Research, Inc., 1977.

Appendix 1A

Annotated Bibliography of Technical Notes and
Reports: Phase III,
Beginning Teacher Evaluation Study

BEGINNING TEACHER EVALUATION STUDY
(BTES)

Phase III-A Technical Reports and Notes

Berliner, David C. Impediments to the study of teacher effectiveness. San Francisco, California: Technical Report 75-11-3, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (29 pages)

This paper outlines problems associated with the independent and dependent variables used in studies of teaching, and related problems in the scientific study of teacher effectiveness. The conclusions reached are that empirical links between teacher behavior and student achievement will be very difficult to establish in the foreseeable future. (This paper, in modified version, is published in the Journal of Research in Science Teaching, 1976, in press; and with their permission, in the Journal of Teacher Education, 1976).

Berliner, D.C. The Beginning Teacher Evaluation Study: Overview and selected findings 1974-1975. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (27 pages)

This paper presents a brief description of major tasks engaged in during the first year of research on this project by the Laboratory. The ethnographic study is discussed in some detail as an example of one research activity of unique interest to educational researchers. (Published as The California Beginning Teacher Evaluation Study: Overview of the Ethnographic Study, by Berliner, D.C. and Tikunoff, W.J., Journal of Teacher Education, 1976).

Berliner, David C., Filby, N.N., Marliave, R.S., Moore, J.E., and Tikunoff, W.J. Experimental teaching units and the identification of a special sample of classrooms for conducting research on teaching. San Francisco, California: Technical Report 76-12-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976. (44 pages plus appendices)

To study the teaching-learning process in classrooms, a small sample of carefully selected teachers at the second and fifth grades were identified for one Phase of the Beginning Teacher Evaluation Study. A unique characteristic of this sample was that it included twenty teachers identified as relatively more effective in influencing achievement gains in their classes and twenty teachers identified as relatively less effective in their ability to teach certain objectives. Procedures and data related to the identification of these two separate groups of teachers from a larger sample are described in this

report. The teachers in the larger sample taught two week units in reading and mathematics, Experimental Teaching Units, and students completed pre and post tests. The smaller sample was selected based on information obtained from the student tests and classroom observations.

Berliner, D.C. and Rosenshine, B. The acquisition of knowledge in the classroom. San Francisco, California: Technical Report IV-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976. (37 pages)

This paper addresses the issue of how learning in classrooms takes place. The effects of different curriculum and different teaching methods in knowledge acquisition is examined, both as main effects and in interaction with student aptitudes. The role of the teacher in communicating the knowledge to be acquired is described. Teachers are viewed as effective in helping students acquire knowledge if they engage in direct instruction. Components of direct instruction are described and data on their relationship to outcomes are presented. (Published in R.C. Anderson, R.J. Spiro, and W.D. Montague (eds) Schooling and the Educational Process, Hillsdale, New Jersey: Earlbaum Associates, 1976, in press).

Erich, Oded, and Shavelson, Richard. The Application of Generalizability Theory to the Study of Teaching. San Francisco, California: Technical Report 76-9-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, September 1976. (150 pages)

This report follows up a study of the stability of teacher behavior over occasions. After showing how classical test theory is inadequate for the problems of measurement in research on teaching, the newer theory of generalizability is presented. An illustrative generalizability study using a completely crossed design, is present. Examples of how variance estimates are obtained for the different design facets are also provided. A procedure is described which helps an investigator to determine the number of observations or raters necessary to determine stable estimates of teacher behavior so that generalizations about that behavior will be possible.

Harnischfeger, Annegret and Wiley, David E. Teaching-learning processes in elementary school: A synoptic view. San Francisco, California: Technical Report 75-3-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (79 pages)

Studies of classroom teaching and learning are criticized for oversimplification. A conceptual framework is proposed, focusing on pupil instructional time in elementary school, exhaustively allocating it to learning pursuits, and tying these to teacher activities and pupil acquisition of knowledge and skill. The conceptual model attempts to establish a linkage between the social and the psychological conditions

of learning, such that the implications of policy and classroom activity changes can be traced to pupil achievement. Teacher planning and implementation emerge as unexamined by vital foci for research on teaching. New policy-relevant school research and teacher evaluation priorities are advocated.

Joyce, Bruce. Variables, designs and instruments in the search for teacher effectiveness. San Francisco, California: Technical Report 75-10-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (117 pages plus appendices)

This report provides a viewpoint about needed instrumentation and design considerations for the study of teaching. The basic questions addressed are: What variables should receive the greatest emphasis in the search for effective teachers? What kinds of designs are available and what are their advantages and disadvantages? And what instruments are available or must be developed to study teacher effectiveness? Discussed in this report are a) the close ties between the teacher and the curriculum materials, making attribution of effects very difficult; b) the ability of students in classrooms to effect teacher behavior as much as teacher behavior effects students. Students have the ability to elicit behavior as well as being the objects of that behavior; and c) the overlap of outcome measures in both reading and mathematics in the elementary grade levels.

Morine, Greta and Vallance, Elizabeth. Special Study B: A study of teacher and pupil perceptions of classroom interaction. San Francisco, California: Technical Report 75-11-6, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (150 pages plus appendices)

Teachers were designated as more and less effective according to their ability to teach specially constructed reading and mathematics curriculum units over two weeks. They were then studied to determine if their perceptions of classroom interaction and interactive decision making were different. They were also studied to determine whether they notice different aspects of teaching when viewing instruction by another teacher. In addition, pupils were also studied to determine what aspects of instruction were salient to them. Four variables related to the teachers thinking about instruction were found worthy of further study. These were the teachers' a) logical/critical thinking, b) comparative thinking, c) negative thinking, and d) amount of input accepted or sought during interactive decision making. In addition, the examination of teacher and pupil perceptions of instruction also yielded variables worthy of further study. In particular, teachers with high pupil gain scores tended to emphasize cognitive aspects of the lesson more, while teachers with low pupil gain scores gave slightly more attention to affective aspects of the lesson. Furthermore, the most salient instructional characteristics for students were cognitive events, not affective events.

Morine, Greta and Vallance, Elizabeth. Special Study C: A study of teacher planning. San Francisco, California: Technical Report 76-3-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976. (150 pages plus appendices)

Twenty more and twenty less effective teachers were studied to determine if they differed in the ways they plan for instruction. The effectiveness of the teachers was determined by their ability to affect gains in reading mathematics over two weeks of instruction with a specially designed curriculum unit. Four variables were noted as distinguishing between the planning of the more and less effective teachers. These were: a) whether the teacher used specific or general information in planning; b) whether the teacher believed students had the potential to learn the material; c) whether the teacher considered the cognitive aspects of instruction; and d) whether the teacher produced much of their own instructional materials.

Shavelson, Richard and Dempsey, Nancy. Generalizability of measures of teacher effectiveness and teaching process. San Francisco, California: Technical Report 75-4-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (72 pages)

This report reviews studies of the stability of measure of teacher effectiveness and behavior. In studying teacher effectiveness over short intervals it appears that teaching the same content to similar students yields moderately stable generalizability coefficients over occasions. However, when teachers teach different content to the same or similar students, the measured effectiveness is not very stable over occasions. Studies of the stability of measured effectiveness over long time intervals yields very low to moderate stability coefficients. When the stability of the behavior of teachers was examined, global ratings appeared to show greater teacher stability over occasions than did frequency counts. Nevertheless, in general, there was very little evidence found to indicate that measures of teacher behavior over time are stable. Instability of teacher behavior seemed to be the common finding. Implications for research on teaching are presented.

Tikunoff, W.J., Berliner, D.C., and Rist, D.C. Special Study A: An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample. San Francisco, California: Technical Report 75-10-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (120 pages plus appendices)

This report describes variables which differentiate the classrooms of more and less effective teachers. Teacher effectiveness was defined as the ability of second and fifth grade teachers to teach specially constructed reading and mathematics curriculum units over two weeks. Twenty more effective and twenty less effective teachers were visited for five days by specially trained ethnographers. Protocols of reading and mathematics instruction were prepared by the ethnographers.

These protocols were rated and variables that distinguished between teachers of differential effectiveness were identified. Twenty-one of these variables were found to distinguish between the more and less effective teachers in their teaching of second grade reading, second grade mathematics, fifth grade reading and fifth grade mathematics. An additional 40 variables also showed some ability to differentiate between more and less effective teachers, but did not generalize across grade levels or subject matters areas. The use of a known sample of teachers varying in effectiveness, and the ethnographic procedures, were both methods found to be very useful in the study of classroom teaching.

Tikunoff, William J., Berliner, David C., and Rist, Ray C. Abstract from Special Study A: An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample. San Francisco, California: Technical Report 75-10-5 (A), Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development. (19 pages)

This document briefly describes a study which is more fully reported in Technical Report 75-10-5, Special Study A: An Ethnographic Study of the Forty Classrooms of the Beginning Teacher Evaluation Study Known Sample. The report describes variables which differentiate the classrooms of more and less effective teachers. Teacher effectiveness was defined as the ability of second and fifth grade teachers to teach specially constructed reading and mathematics curriculum units over two weeks. Twenty more effective and twenty less effective teachers were visited for five days by specially trained ethnographers. Protocols of reading and mathematics instruction were prepared by the ethnographers. These protocols were rated and variables that distinguished between teachers of differential effectiveness were identified. Twenty-one of these variables were found to distinguish between the more and less effective teachers in their teaching of second grade reading, second grade mathematics, fifth grade reading and fifth grade mathematics. An additional 40 variables also showed some ability to differentiate between more and less effective teachers, but did not generalize across grade levels or subject matter areas. The use of a known sample of teachers varying in effectiveness, and the ethnographic procedures, were both methods found to be very useful in the study of classroom teaching.

Tikunoff, William J., Berliner, David C., and Rist, Ray C. Sample of second grade classroom protocols from Special Study A of the Beginning Teacher Evaluation Study. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, March-April, 1975. (13 pages plus appendices)

Sample of Second Grade Classroom Protocols collected within this volume are examples of the protocols from grade 2 developed by the ethnographers associated with Special Study A: "An Ethnographic Study of the Forty Classrooms of the Beginning Teacher Evaluation Study". Twenty teachers at both the second and fifth grades - forty teachers in all - were observed for one week by an ethnographer who observed

the teaching of reading and mathematics. Ethnographer field notes were dictated and transcribed into formal protocols. In addition, informal protocols were developed from informal observations such as conversations with teachers or students. The ethnographers also developed a summary profile for each classroom based on Jules Henry's "A Cross-Cultural Outline of Education". Protocols in this volume were selected from several classrooms.

Tikunoff, William J., Berliner, David C., and Rist, Ray C. Sample of fifth grade classroom protocols from Special Study A of the Beginning Teacher Evaluation Study. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, March-April, 1975. (12 pages plus appendices)

Sample of Fifth Grade Classroom Protocols collected within this volume are examples of the protocols from grade 5 developed by the ethnographers associated with Special Study A: "An Ethnographic Study of the Forty Classrooms of the Beginning Teacher Evaluation Study". Twenty teachers at both the second and fifth grades - forty teachers in all - were observed for one week by an ethnographer who observed the teaching of reading and mathematics. Ethnographer field notes were dictated and transcribed into formal protocols. In addition, informal protocols were developed from informal observations such as conversations with teachers or students. The ethnographers also developed a summary profile for each classroom based on Jules Henry's "A Cross-Cultural Outline of Education". Protocols in this volume were selected from several classrooms.

Experimental Teaching Unit: Second Grade Mathematics. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1974. (141 pages)

The Experimental Teaching Unit is designed to provide common instructional objectives, in an area which is educationally important but not commonly covered within the curriculum to provide some consistency in instruction across a number of classes, while also providing sufficient flexibility to enable teachers to maintain their normal teaching styles. The Units contain instructional objectives and a number of teaching activities, as well as student achievement tests.

The purpose of this unit is to develop basic concepts of measurement. The child will have a variety of experiences which make measurement concepts meaningful to him. This unit will provide some review of counting skills and simple addition and subtraction skills. The primary focus, however, is to build concepts of measurement in length, weight or mass, volume or capacity, and area and extend them into a simple introduction to the metric system. Activities in each subtopic of this Experimental Teaching Unit suggest ways to accomplish that. The subtopics are: comparisons, ordering, conservation, arbitrary units of measuring, and standard units of measuring.

Experimental Teaching Unit: Second Grade Reading. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1974. (191 pages)

The Experimental Teaching Unit is designed to provide common instructional objectives, in an area which is educationally important but not commonly covered within the curriculum to provide some consistency in instruction across a number of classes, while also providing sufficient flexibility to enable teachers to maintain their normal teaching styles. The Units contain instructional objectives and a number of teaching activities, as well as student achievement tests.

This unit deals with important reading skills of following directions, understanding and following a sequence of actions, word structure and prefixes and suffixes.

Experimental Teaching Unit: Fifth Grade Mathematics. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1974. (319 pages)

The Experimental Teaching Unit is designed to provide common instructional objectives, in an area which is educationally important but not commonly covered within the curriculum to provide some consistency in instruction across a number of classes, while also providing sufficient flexibility to enable teachers to maintain their normal teaching styles. The Units contain instructional objectives and a number of teaching activities, as well as student achievement tests.

This unit deals with the concept of probability, the use of probability to make predictions, and procedures to gather and display probability data using graphs. Fractions are used to express probabilities.

Experimental Teaching Unit: Fifth Grade Reading. San Francisco, California: Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1974. (225 pages)

The Experimental Teaching Unit is designed to provide common instructional objectives, in an area which is educationally important but not commonly covered within the curriculum to provide some consistency in instruction across a number of classes, while also providing sufficient flexibility to enable teachers to maintain their normal teaching styles. The Units contain instructional objectives and a number of teaching activities, as well as student achievement tests.

This unit is designed to improve a student's reading comprehension by increasing his understanding of the way language works. To improve reading for meaning, the student is introduced to the concept

of paraphrase. A particular idea can be stated in a number of different ways. The external form of the message can change as sentences are rewritten, but the underlying meaning stays the same. This concept of paraphrase is applied at three levels: (1) different words can be used to represent the same concept (synonyms); (2) sentences can be rewritten to express the same idea; and (3) paragraphs or stories have a major topic and can contain ideas written in different ways. Another section of the unit gives the student practice in using context clues. As a skilled reader processes a sentence, he uses his knowledge of language to make predictions about what words will appear. By integrating words into the larger unit of the sentence or paragraph, the reader can decode words with a minimum of visual cues. He relies on the redundancy of language to fill in additional features.

BEGINNING TEACHER EVALUATION STUDY
(BTES)

PHASE III-A CONTINUATION

Marliave, Richard S. A review of the findings of Phase II. San Francisco, California: Technical Note I-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, May, 1976. (33 pages)

This document examines the implications of the Phase II report specifically for observation variables that might be used in future Beginning Teacher Evaluation Study (BTES) research, focusing on those variables and patterns of variables that showed the strongest and most consistent relationships with student learning the Phase II report. The variables are discussed within four basic categories: 1) teacher behaviors, 2) student behaviors, 3) teacher characteristics, and 4) student characteristics. This document is written with the assumption that the reader is familiar with the design and methodology of the Phase II study. The Phase II report examines the data collection, data analysis, and conclusions of one of the most comprehensive and complex studies of teaching conducted to date.

Marliave, Richard S. Observable classroom variables. San Francisco, California: Technical Report I-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, July, 1976. (28 pages)

This report provides a brief review of BTES documents from Phase II and Phase III-A, including responses from nationally-known researchers suggesting variables for future BTES research. Seven basic observable classroom variable domains have been found to receive support from several of the BTES sources reviewed. These seven variable domains are: 1) the quantity of academic instruction received by the student; 2) the strength of the academic orientation of instruction; 3) the extent to which instruction is tailored to the needs of individual students; 4) the clarity of instructional communication; 5) the teacher's instructional involvement and contact with students; 6) student attention to the academic task; and 7) the existence of a positive, family-like classroom atmosphere. It was also concluded that variables related to information processing by the teacher make up a domain of non-observable classroom variables. In addition, several methodological implications of the research reviewed were discussed.

Marliave, Richard. Findings of clinical field observations. San Francisco, California: Technical Note I-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, July, 1976. (21 pages)

This document describes the clinical observation activities in a sample

of California classrooms as part of the developmental work for Phase III of the Beginning Teacher Evaluation Study. The actual observation procedures are documented and the conclusions reached on the basis of those observations are presented. Conclusions related to time-based observation of student behavior and teacher-student interaction, as well as related data collection procedures, are presented. Additional conclusions relate to the usability of rating scales, questionnaires, and interviews.

Marliave, Richard S., Fisher, Charles W., Filby, Nikola N., Dishaw, M. The Development of instrumentation for a field study of teaching. San Francisco, California: Technical Report I-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, February, 1977. (43 pages plus appendices)

This report describes time coding forms, rating scales, interview schedules, and a questionnaire, all of which are instruments designed for use in Phase III-B of BTES. The BTES developmental activities leading to the design of this instrumentation are reviewed. Design and field testing activities relevant to each instrument are also presented. Inter-observer reliability data obtained during the field test of the time coding forms are presented and discussed. In addition, all of the forms, scales, and manuals that comprise these instruments are presented in the body and the appendices of this report.

Design for a study of instructional time. San Francisco, California: Technical Note II-1, Continuation of Phase III-A, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, August 15, 1975. (34 pages)

This report describes the research design for a study of instructional time, one of the research efforts of Phase III-A of the Beginning Teacher Evaluation Study. The variables to be studied, the sample, the data collection and analysis procedures are described, and a research time schedule is included. The instruments to be used in the research are described, and a research time schedule is included. The instruments to be used in the research are described in Technical Note II-2, Instruments for the Study of Instructional Time. The results of this research effort are presented in several companion volumes, Technical Report II-3, A Study of Instructional Time in Grade 2 Mathematics; Technical Report II-4, A study of Instructional Time in Grade 2 Reading; and Technical Report II-5, Instructional Time Allocation in Fifth Grade Reading.

Instruments for the study of instructional time. San Francisco, California: Technical Note II-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, August 15, 1975. (38 pages)

This document describes the instruments, and the development of the instruments, for recording of teacher time allocation (teacher logs)

and classroom observation for the study of instructional time in Phase III-A of the Beginning Teacher Evaluation Study (BTES). The research design for the study is presented in Technical Note II-1, Design for a Study of Instructional Time. The data collection procedures and instruments are the result of research efforts conducted in earlier phases of the BTES and this development is chronicled in this Technical Note.

Harnischfeger, Annegret and Wiley, David E. Notes for a field study of instructional time. San Francisco, California: Technical Report 75-10-11, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, October, 1975. (33 pages)

This paper follows up on an early report which presented a theory on instructional processes that relied on a time metric. In this paper a design for field studied within this theoretical orientation is presented. Methods to obtain information on how teachers allocate time so that grouping and individualization strategies can be studied are discussed. Sampling procedures for observations are recommended, and a data file structure for this kind of information is explained.

Fisher, Charles W., Marliave, Richard S., Filby, Nikola N., Cahen, Leonard S., Moore, Jeffrey E., and Berliner, David C. A study of instructional time in grade 2 mathematics. San Francisco, California: Technical Report II-3, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976. (159 pages)

This report contains the results of a field study of the relationship between instructional time and student achievement. Quantities of time allocated to mathematics content categories and instructional settings are described for a forty-day instructional period. Descriptions of student "engaged time" for a ten-day period are also included. The data were collected in six grade 2 classrooms. Relationships between both allocated time and engaged time in specific mathematics content categories and achievement in those content categories were investigated. Large differences were found in the amounts of time allocated to mathematics content categories. Large differences were also found in the amounts of time students were engaged in instructional activities. Positive relationships were found between amount of instruction and student achievement.

Fisher, Charles W., Filby, Nikola N., Marliave, Richard S., Cahen, Leonard S., Moore, Jeffrey E., and Berliner, David C. A study of instructional time in grade 2 reading. San Francisco, California: Technical Report II-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976. (133 pages plus appendix)

This report contains the results of a field study of the relationship between instructional time and student achievement. Quantities of time allocated to reading categories and instructional settings are

described for a forty-day instructional period. Descriptions of student "engaged time" for a ten-day period are also included. The data were collected in nine grade 2 classrooms. Relationships between both allocated time and engaged time in specific reading content categories and achievement in those content categories were investigated. Large differences were also found in the amounts of time students were engaged in instructional activities. Positive but relatively weak relationships were found between amount of instruction and student achievement.

Instructional time allocation in fifth grade reading. San Francisco, California: Technical Report II-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, December, 1976. (53 pages plus appendices)

This report describes the naturally occurring variations in time allocated to reading instruction in a sample of grade five classes selected as part of the Beginning Teacher Evaluation Study. Students completed a set of reading tests in early and late fall and teachers maintained records of instructional time allocated, setting, and content covered, for the students. Ceiling effects in the reading tests precluded an analysis of the relationship of time to achievement. The time data were analyzed to identify variations across classes and differences among students within a class in terms of allocated time to instructional settings and content areas.

Filby, Nikola N. and Dishaw, Marilyn. Development and refinement of reading and mathematics tests for grades 2 and 5. San Francisco, California: Technical Report III-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975. (213 pages, plus Appendix C, mathematics items, and Appendix D, reading items, which are bound separately)

This report describes initial and future plans to develop tests that are maximally sensitive to instruction. Two important issues to address are the content coverage of the tests (opportunity to learn) and the reactivity of the measure to instruction. The ability of students must be minimally related to change in performance, and when the opportunity-to-learn is high, that factor must be maximally related to performance in order to have a test that is useful in studies of teaching. An ability by opportunity-to-learn matrix is constructed for every item in the battery of reading and mathematics tests used in second and fifth grade in order to develop as sensitive a test as possible for the study of teaching. These procedures are described in this report.

Design for use of individual testing in the study of instructional time. San Francisco, California: Technical Note III-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, September, 1975. (12 pages)

This report describes the general plan for a study to clarify the

interpretation of group performance measures of student achievement. Group tests are the most efficient procedure for collecting achievement information from a large number of students, but they also require item formats which may be different from the form in which students normally demonstrate or apply their learning. To provide information about some of these potential differences, to be used in the interpretation of group test scores, two studies are planned, one of second grade students in the areas of reading and mathematics, and one of second and fifth grade students in reading comprehension.

Instruments for Individual Assessment of Achievement. San Francisco, California: Technical Note III-3, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, September 15, 1975. (15 pages)

One task of the Beginning Teacher Evaluation Study was to develop and refine student achievement tests. Individual testing was conducted to provide input information relative to the group achievement tests. This technical note describes the materials and the procedures used for the individual student testing.

Filby, Nikola N. and Dishaw, Marilyn M. Construct validation of group-administered achievement tests through individual testing. San Francisco, California: Technical Note III-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, April, 1977. (62 pages plus appendices)

This report describes part of the work done to develop student achievement tests in reading and mathematics for use in Phase III-B of the Beginning Teacher Evaluation Study. While the achievement battery consists primarily of group-administered multiple-choice test items, several individual-testing studies were conducted to examine the validity of group-test measures and to gain insight for the improvement of group-test items. The sample of students, data collection and analysis procedures, and results and conclusions are presented. The actual revisions in the test battery which resulted from this effort are described in Technical Report III-6, Refinement of Reading and Mathematics tests through an Analyses of Reactivity.

Filby, Nikola N. Progress report on reactivity analyses (October - December test data). San Francisco, California: Technical Note III-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, April 30, 1976. (10 pages)

This report summarizes the procedures used to collect and analyze data to evaluate the degree to which test items are reactive to instruction. This progress report describes actual steps conducted to complete the study described in Technical Report III-1, Development and Refinement of Reading and Mathematics Tests for Grades 2 and 5 of the Beginning Teacher Evaluation Study. The results of these analyses are described in Technical Report III-6, Refinement of Reading and Mathematics Tests

through an Analysis of Reactivity.

Filby, Nikola N. and Dishaw, Marilyn. Refinement of reading and mathematics tests through an analysis of reactivity. San Francisco, California: Technical Report III-6, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976. (57 pages plus appendices)

This report follows up Technical Report III-1 and describes major analyses of test reactivity to instruction. Year-long teacher records were used to determine opportunity-to-learn. An ability by opportunity-to-learn matrix is reported for each subtest. Item reactivity is also reported. Test revisions, as dictated by reactivity analyses and other pilot test data, are summarized. The report compares reactivity analysis to standard psychometrics as a method of producing tests which are sensitive measures of classroom learning.

Marliave, Richard, Cahen, Leonard S., and Berliner, David C. Prolegomenon on the concept of appropriateness of instruction. San Francisco, California: Technical Report IV-1, Phase III-A Continuation, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, July, 1977. (14 pages)

This report discusses the question of the appropriateness of teacher behaviors in relation to both the counting of frequencies of teacher behaviors and the interpretation of the impact of teacher behaviors on student learning. The recent focus, in research on teaching, on the counting of teacher behaviors gives rise to a concern for the meaning, as well as the frequency, of teacher behaviors. The report describes three different definitions of appropriateness and the related systems for measuring teacher behaviors. The first definition considers appropriateness in terms of teacher decision making; the second in terms of teacher behavior in classroom settings; the third in relation to Academic Learning Time.

Filby, Nikola N. Plans for further development of experimental teaching units. San Francisco, California: Technical Note V-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, March, 1976. (7 pages)

The report describes the problems related to control of or variety in curriculum in studies of teaching and the procedures used in one Phase of the Beginning Teacher Evaluation Study to control to some extent the curriculum in reading and mathematics. The major focus of the document is the plan for further study of aspects of Experimental Teaching Units designed to provide varying degrees of control of curriculum to facilitate research on teaching.

Filby, Nikola N. Theoretical and empirical consideration of design issues related to experimental teaching units. San Francisco, California: Technical Note V-3, Phase III-A Continuation, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, June, 1977. (13 pages)

This Technical Note discusses some theoretical and empirical issues that influenced the selection of a research design for the major field study carried out in Phase III-B of BTES. One line of research in the hypothesis-generation phase of the study used an Experimental Teaching Unit (ETU) to create a semi-standardized situation for comparison of teachers. Another line of research focused on naturally-occurring behavior, particularly instructional time. In the instructional time framework, a different research design was used, involving an "objectives-only" ETU. As plans were made for the major field study, design issues were reconsidered. Technical Note V-2, plans for Further Development of Experimental Teaching Units, stated the need to reconsider the reasons for an experimental teaching unit and to evaluate the alternative design approaches that had been used thus far. This Technical Note reviews and discusses issues of research design, presents data from alternative approaches, and draws conclusions for the Phase III-B design.

BEGINNING TEACHER EVALUATION STUDY
(BTES)

PHASE III-B

Howell, W. Pearl and Rice, Carol F. Field Coordination Phase III-B, Beginning Teacher Evaluation Study. San Francisco, California: Technical Note II-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, July, 1977. (23 pages plus appendices)

This technical report gives an overview of coordination activities that occurred during the data collection for Phase III-B of the Beginning Teacher Evaluation Study. A history of the recruitment, selection, and training of field personnel, a description of coordination efforts with teachers and school districts, a discussion of the various aspects of scheduling, and a discussion of procedures for logging-in and checking the quality of the field data are presented. Each major section of the report summarizes in sequence the tasks required for collection of the field data and provides supporting documents. Other tasks during the data collection phase, such as log training for teachers, instrumentation development, and observation training are presented in other BTES technical notes and reports.

Cahen, Leonard S. The development of a selection battery for Phase III-B. San Francisco, California: Technical Note III-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, November 30, 1976. (18 pages)

This report describes the development of a Selection Battery for choosing target students for Phase III-B of the Beginning Teacher Evaluation Study. A discussion of the reasons for selecting target students for intensive classroom observation precedes the description of procedures used to select from the larger test battery a sample of test items to form the Selection Battery. A second report (Technical Note III-1, Part II) describes the procedures actually used for selecting target students from among the students to whom the Selection Battery was administered.

Cahen, Leonard S. Selection of second and fifth grade target students for Phase III-B. San Francisco, California: Technical Note III-1 (Part II), Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, January, 1977. (18 pages)

The actual selection of target students from second and fifth grade classrooms, from the group of students to whom the Selection Battery was administered, is described in this report, including the statistical procedures and cut-off scores. A companion report (Technical Note III-1, The Development of a Selection Battery for Phase III-B, Leonard S. Cahen) describes the procedures for developing

the Selection Battery from the full battery of student achievement tests, as well as the reasons for using a sample of target students, selected on the basis of achievement, for intensive classroom observation as part of the data collection for one phase of the Beginning Teacher Evaluation Study.

Cahen, Leonard S. Report on "A" testing. San Francisco, California: Technical Note III-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, December 23, 1976. (8 pages)

This document describes the "A" testing, Phase III-B of the Beginning Teacher Evaluation Study, which took place in October, 1976. A sample of target students, 6 per class, in a sample of second and fifth grade classes in California completed the achievement battery in reading and mathematics. Administration procedures, results, and psychometric properties of the tests are described.

Cahen, Leonard S. Report on "B" testing. San Francisco, California: Technical Note III-3, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, February 25, 1977. (9 pages)

This document describes the "B" testing, Phase III-B of the Beginning Teacher Evaluation Study, which took place in December 1976. A sample of target students, 6 per class, in a sample of second and fifth grade classes in California completed the achievement battery in reading and mathematics. Administration procedures, results, and psychometric properties of the tests are described. This is a companion document to Technical Note III-2, Report of the "A" Testing.

Cahen, Leonard S. Report of "C" testing. San Francisco, California: Technical Note III-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, June, 1977. (11 pages)

This document describes the "C" testing, Phase III-B of the Beginning Teacher Evaluation Study, which took place in May, 1977. target students, 6 per class, in a sample of second and fifth grade classes in California completed the achievement battery in reading and mathematics. Administration procedures, results, and psychometric properties of the tests are described. This is a companion document to Technical Note III-2, Report of the "A" and "B" Testings.

Cahen, Leonard S. Report on "D" testing. San Francisco, California: Technical Note III-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, October, 1977. (11 pages)

This document describes the "D" testing, Phase III-B of the Beginning Teacher Evaluation Study, which took place in September, 1977. A sample of target students six per class, in a sample of second and fifth grade classes in California completed the achievement battery in

reading and mathematics. Administration procedures results, and psychometric properties of the tests are described. This is a companion document to Technical Note III-2, Report of the "A" Testing, Technical Note III-3, Report of the "B" Testing, and Technical Note III-4, Report of the "C" Testing.

Filby, Nikola N. and Marliave, Richard. Descriptions of distributions of ALT within and across classes during the A-B period. San Francisco, California: Technical Note IV-1a, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, September, 1977. (44 pages)

This report provides descriptive statistical information from Phase III-B of the Beginning Teacher Evaluation Study, being conducted by the Commission for Teacher Preparation and Licensing. With the sample of second and fifth grade classrooms in public schools in California, approximately six students were identified as target students to be studied during the school year. These students were observed approximately weekly during the period from October to December and again during the period from January to May. The observers recorded student time within subject matter areas of reading and mathematics, focusing on the content of student instructional tasks. ALT (Academic Learning Time) is defined as time on tasks of intermediate difficulty for the student directly related to the measures of achievement under study during which a student is engaged in the instructional task. The report provides descriptive information on ALT for individuals within classes and comparisons across classes for the period from October to December during the 1976-77 school year.

Fisher, Charles W., Filby, Nikola N., and Marliave, Richard S. Descriptions of distributions of ALT within and across classes during the B-C period. San Francisco, California: Technical Note IV-1b, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, October, 1977. (49 pages)

This report provides descriptive statistical information from Phase III-B of the Beginning Teacher Evaluation Study, being conducted by the Commission for Teacher Preparation and Licensing. Within the sample of second and fifth grade teachers in classrooms in public schools in California, approximately six students were identified as target students to be studied during the school year. These students were observed approximately weekly during the period from January to May. The observers recorded student time within subject matter areas of reading and mathematics, focusing on the content of student instructional tasks, the degree of difficulty of instructional tasks for individual students, and student engagement on those tasks. ALT (Academic Learning Time) is defined as time on tasks of intermediate difficulty for the student, directly related to the measures of achievement under study, during which a student is engaged in the instructional task. The report provides descriptive information on ALT for individuals within classes and comparisons across classes for the period from January to May during the 1976-77 school year.

Dishaw, Marilyn. Descriptions of allocated time to content areas for the A-B period. San Francisco, California: Technical Note IV-2a, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, July, 1977. (42 pages plus appendices)

This report provides descriptive data on allocated time in reading and mathematics for grades 2 and 5 as reported by a sample of California teachers participating in one Phase of the Beginning Teacher Evaluation Study. The procedures used by teachers to record information are described and data are presented for the A-B period, October to December. Class means are provided on the amount of time accumulated over the A-B period in specific content categories. Daily means and the mean percent of total time allocated to specific content areas are also presented, along with teacher estimates of task difficulty. A companion report (Technical Note IV-2b, Descriptions of Allocated Time to Content Areas for the B-C Period, Marilyn Dishaw) presents comparable data for the school year from January to May (1976-77).

Dishaw, Marilyn. Descriptions of allocated time to content areas for B-C period. San Francisco, California: Technical Note IV-2b, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, August, 1977. (31 pages)

This report describes the allocation of time in reading and mathematics to specific content categories in a sample of second and fifth grade classes, as reported by teachers participating in one Phase of the Beginning Teacher Evaluation Study. Class means are provided in the amount of time accumulated over the period from January to May. Daily means and the mean percent of total time allocated to specific content areas are also presented, along with teacher estimates of task difficulty. A companion report (Technical Note IV-2a, Descriptions of Allocated Time to Content Areas for the A-B Period, Marilyn Dishaw) describes the data collection procedures and presents data from October to December of the same school year (1976-77).

Filby, Nikola N. Description of patterns of teaching behavior within and across classes during the A-B period. San Francisco, California: Technical Note IV-3a, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, October, 1977. (54 pages)

This document provides a description of the data collection procedures and the results of data collection for observation of teaching behaviors in the sample classrooms in Phase III-B of the Beginning Teachers Evaluation Study, being conducted by the Far West Laboratory for the Commission for Teacher Preparation and Licensing. During the 1976-77 school year, the behaviors of the teachers within the Phase III-B sample were observed approximately once each week for the entire school day, with the observations focusing on the teaching of reading and mathematics. A sample of approximately six students were selected in each classroom as the focus of observation. Any behaviors of teachers directed toward these students, either individually or as members

of groups, were recorded by the observers. The data were summed for the period from October to December of the school year, covering approximately seven weeks of observations, and are presented by class as well as summed for all classes within the study.

Filby, Nikola, N., and Fisher, Charles W. Description of patterns of teaching behavior within and across classes during the B-C period. San Francisco, California: Technical Note IV-3b, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, October, 1977. (54 pages)

This document provides a description of the data collection procedures and the results of data collection for observation of teaching behaviors in the sample classrooms in Phase III-B of the Beginning Teacher Evaluation Study, being conducted by the Far West Laboratory for the Commission for Teacher Preparation and Licensing. During the 1976-77 school year, the behaviors of the teachers within the Phase III-B sample were observed approximately once each week for the entire school day, with the observations focusing on the teaching of reading and mathematics. A sample of approximately six students, either individually or as members of groups, were recorded by the observers. The data were summed for the period from January to May of the school year, covering approximately 15 weeks of observations, and are presented by class as well as summed for all classes within the study.

Marliave, Richard S., Fisher, Charles W., and Dishaw, Marilyn M. Academic learning time and student achievement in the A-B period. San Francisco, California: Technical Note V-1a, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1977. (135 pages)

This document describes the analysis procedures and presents results for the relationship between ALT and student achievement. Analyses are restricted to the A-B period. Achievement is regressed on entering achievement level and allocated time, student engagement rates and student error rates. Results are presented for analysis based on individual students scores and for class mean scores.

Marliave, Richard S., Fisher, Charles W., and Dishaw, Marilyn M. Academic learning time and student achievement in the B-C period. San Francisco, California: Technical Note V-2a, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1978. (129 pages)

This document contains analyses of ALT and achievement relationships based on the B-C data from Phase III-B of BTES. It parallels Technical Note V-1a, but for a later and longer time period.

Filby, Nikola N., and Cahen, Leonard S. Teaching behavior and academic learning time in the A-B period. San Francisco, California: Technical Note V-1b, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1977. (82 pages plus appendices)

This document presents analyses of teaching behaviors and facets of academic learning time. The data were collected in samples of grade 2 and grade 5 classrooms by intensive observation, interview and rating procedures. Teaching functions such as diagnosis, prescription, presentation, monitoring and feedback are related to student engagement and student error rate. The document also presents analysis of classroom context variables.

Filby, Nikola N., and Cahen, Leonard S. Teaching behavior and academic learning time in the B-C period. San Francisco, California: Technical Note V-2b, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1978. (91 pages plus appendices)

This report parallels Technical Note V-1b, but covers analyses of data collected during the longer B-C period.

Berliner, David C., Filby, Nikola N., Marliave, Richard S., Weir, Christina D., Fisher, Charles W. Description of classes and plan for an intervention study. San Francisco, California: Technical Note VI-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, December, 1976. (31 pages)

This technical note contains sections on the philosophy of the intervention study, the procedures and methodology to be used, the history of selection of classes, the instructional model, and examples of possible interventions in classes selected for inclusion in the intervention study in Phase III-B of the Beginning Teacher Evaluation Study. A final section, titled Summary Comments, contains the three major recommendations of the Commission's Research Advisory Board who reviewed this document. The study focuses on the process of change in teacher and student behaviors in classrooms.

Berliner, David C., Filby, Nikola N., Marliave, Richard S., and Weir, Christina D. An intervention in classrooms using the Beginning Teacher Evaluation Study model of instruction. San Francisco, California: Technical Report VI-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1978. (228 pages)

During Phase III-B of BTES, the ALT model of classroom instruction was used to guide interventions in grade 2 classrooms. In four classrooms intensive interaction occurred between researchers and teachers in an attempt to implement procedures which would increase the amount of academic learning time accumulated by students. An additional group of teachers participated in workshops dealing with the ALT model. The

study was carried out in a time-series design framework. This report covers the procedures used, case studies of individual classes and results of the interventions.

Fisher, Charles W., and Berliner, David C. Quasi-clinical inquiry in research on classroom teaching and learning. San Francisco, California: Technical Report VI-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1977. (25 pages)

This report critically examines traditional research on teaching and learning. Advantages and disadvantages of frequently-used methodology are discussed. The paper concludes with proposals for a more clinically oriented methodology. A portion of the report deals with the relationship between research and practice in education and discusses the impact of traditional and quasi-clinical research styles on this relationship.

Fisher, Charles W., Filby, Nikola N., Marliave, Richard S., Cahen, Leonard S., Dishaw, Marilyn M., Moore, Jeffrey E., and Berliner, David C. Teaching behaviors, academic learning time and student achievement: Final report of Phase III-B, Beginning Teacher Evaluation Study. San Francisco, California: Technical Report-V-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1978. (346 pages plus appendices)

This document is the final report and technical summary of Phase III-B of BTES. The report draws upon the technical notes and reports listed above. The report contains an overview of Phase III of BTES, a description of the ALT model of classroom learning, description of the data collection procedures, results for the relationship between ALT and achievement, results for the relationship of teaching processes to student engagement and student error rate, analyses of the overall ALT model, qualitative information on "outlier" classes, results for the relationship of student attitudes to academic learning time, and an analysis of student achievement retention. The final portion of the report presents a summary of the findings and the conclusions drawn from the study.

Fisher, Charles W., Berliner, David C., Filby, Nikola N., Marliave, Richard S., Cahen, Leonard S., Dishaw, Marilyn M., and Moore, Jeffrey E. Teaching and learning in the elementary school: A summary of the Beginning Teacher Evaluation Study. San Francisco, California: Technical Report VII-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, June 1978.

This document summarizes Phase III of the Beginning Teacher Evaluation Study. This report is intended for a general audience. The context of the study, the major findings, conclusions, and recommendations are discussed.

Appendices 3A through 3I

APPENDIX	TITLE	PAGES
3A	Descriptive Information on Teachers and Target Students	3A-1 to 3A-2
3B	Descriptive Information on Socioeconomic Status for Target and Non-Target Students	3B-1 to 3B-4
3C	Psychometric Data on Achievement Tests and Attitude Scales	3C-1 to 3C-25
3D	Teacher Log Content Categories	3D-1 to 3D-4
3E	Brief Description of Observation Categories	3E-1 to 3E-8
3F	Inter-observer Reliability Coefficients for Observation Categories	3F-1 to 3F-4
3G	Correspondence among Observation, Congruency, and Teacher Log Content Categories	3G-1 to 3G-4
3H	Reliability of Observed Engagement Variables	3H-1 to 3H-6
3I	Reliability of Observed Student Error Rates	3I-1 to 3I-6

Table 3A-1
Sex and Ethnicity of Teachers

Grade 2

	Asian	Black	Chicano	White	Total
Male:	1	0	2	3	6
Female:	1	4	0	16	21

Grade 5

	Asian	Black	Chicano	White	Total
Male:	0	0	1	7	8
Female:	1	2	0	17	20

Note: There were more teachers than classes in the sample because some classes were team-taught.

Table 3A-2

Sex and Ethnicity of Target Students

Grade 2

	Asian	Black	Chicano	White	Total
Male:	6	12	16	31	65
Female:	3	14	4	50	71

Grade 5

	Asian	Black	Chicano	White	Native American	Total
Male:	6	4	7	49	1	67
Female:	4	5	8	41	0	58

Table 3B-1

Percentage of Second Grade Students Classified
by Parental Occupation Groups

Classroom Code	N	GROUP* A	GROUP** B	GROUP*** C	GROUP**** D
01	24.0	5.0	25.0	65.0	5.0
02	26.0	25.0	30.0	30.0	15.0
03	27.0	2.0	20.0	48.0	30.0
04	27.0	20.0	40.0	40.0	0.0
05	25.0	4.0	16.0	80.0	0.
06	28.0	10.0	3.0	50.0	37.0
07	21.0	10.0	25.0	60.0	5.0
08	22.0	25.0	25.0	50.0	0.0
09	30.0	0.0	25.0	50.0	25.0
10	28.0	20.0	10.0	30.0	40.0
11	16.0	0.0	37.0	41.0	22.0
12	17.0	6.0	15.0	64.0	15.0
13	22.0	1.0	44.0	50.0	5.0
15	34.0	10.0	20.0	50.0	20.0
16	35.0	6.0	50.0	40.0	4.0
17	18.0	5.0	20.0	70.0	5.0
18	30.0	0.0	80.0	20.0	0.0
19	21.0	5.0	10.0	80.0	5.0
20	22.0	0.0	15.0	45.0	40.0
21	14.0	4.0	10.0	50.0	36.0
22	27.0	3.0	27.0	50.0	20.0
23	22.0	5.0	20.0	60.0	15.0
24	28.0	5.0	30.0	30.0	35.0
25	24.0	40.0	35.0	5.0	20.0

* Executive/Professional/Managers

** Semi-Professional/Clerical/Sales Workers/Technicians

*** Skilled and Semi-Skilled Employees

**** Unskilled Employees (including those on welfare and those not regularly employed.)

Table 3B-2

Number and Percentages of Second Grade Target Students
by Parental Occupation Classification Groups

	GROUP* A	GROUP** B	GROUP*** C	GROUP**** D	Total
Target N Students Reported	13	42	68	14	137
Target Students % Reported	9.5	30.7	49.6	10.2	
% All Students	9.0	27.4	47.0	16.6	

* Executive/Professionals/Managers

** Semi-Professional/Clerical/Sales Workers/Technicians

*** Skilled and Semi-Skilled Employees

**** Unskilled Employees (including those on welfare
and those not regularly employed.)

Table 3B-3
 Percentage of Fifth Grade Students Classified
 by Parental Occupation Groups

Classroom Code	N	GROUP* A	GROUP** B	GROUP*** C	GROUP**** D
01	31	15	10	65	10
03	26	0	5	95	0
04	28	40	40	20	0
05	30	6	10	18	66
06	20	4	5	76	15
08	14	4	32	60	4
09	30	5	15	70	10
10	29	10	40	40	10
11	21	19	23	23	35
12	20	0	49	51	0
13	29	(N O T R E P O R T E D)			
14	25	3	23	71	3
16	24	0	34	66	0
17	27	16	28	44	12
18	17	5	40	35	20
19	17	0	25	50	25
21a	27	10	80	10	0
22	26	10	40	40	10
23	24	50	40	10	0
24	23	15	40	40	5
25	17	(N O T R E P O R T E D)			
26	30	15	20	60	5
27	26	10	35	35	20

* Executive/Professionals/Managers

** Semi-Professional/Clerical/Sales Workers/Technicians

*** Skilled and Semi-Skilled Employees

**** Unskilled Employees (including those on welfare and those not regularly employed.)

Table 3B-4

Number and Percentages of Fifth Grade Target Students
by Parental Occupation Classification Groups

	GROUP* A	GROUP** B	GROUP*** C	GROUP**** C	Total
Target N Students Reported	17	41	55	9	122
Target Students % Reported	13.9	33.6	45.1	7.4	
% All Students	11.9	29.8	46.4	12.0	

* Executive/Professional/Managers

** Semi-Professional/Clerical/Sales Workers/Technicians

*** Skilled and Semi-Skilled Employees

**** Unskilled Employees (including those on welfare and those not regularly employed.)

Table 3C-1
Psychometric Data
Grade 2 Reading Subtests
"A" Administration
(N = 150)

Subtests	No. of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent Passing 80% or more of Items in Subtest
Decoding Consonant Blends	12	6.79	3.63	.57	.69	1.94	23
Decoding Variant Consonants	14	4.41	5.60	.32	.72	2.83	15
Decoding Long Vowels	20	7.89	6.05	.40	.78	2.79	10
Decoding Complex Patterns	16	4.86	4.28	.30	.69	2.30	1
Compound Words	8	4.71	2.26	.59	.77	1.01	25
Root Words and Affixes - D	9	3.87	1.81	.43	.53	1.17	0
Root Words and Affixes - C	8	1.59	2.83	.20	.60	1.68	1
Syllables	10	2.43	2.68	.24	.36	2.04	0
Synonyms	16	4.36	4.48	.27	.77	2.06	1
Pronoun Reference	10	1.91	2.74	.19	.41	1.99	0

Table 3C-2

Psychometric Data
Grade 2 Reading Subtests
"B" Administration
(N = 146)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Decoding Consonant Blends	12	8.26	3.17	.69	.64	1.78	40
Decoding Variant Consonants	14	5.47	5.42	.39	.73	2.73	18
Decoding Long Vowels	20	10.60	6.76	.53	.82	2.60	21
Decoding Complex Patterns	16	6.20	4.62	.39	.75	2.26	12
Compound Words	8	5.51	2.09	.69	.76	0.96	40
Root Words and Affixes - D	9	4.74	1.84	.53	.52	1.20	9
Root Words and Affixes - C	8	2.91	2.70	.36	.55	1.71	16
Syllables	10	2.93	2.69	.29	.40	1.97	4
Synonyms	16	5.13	4.96	.32	.81	2.07	6
Pronoun Reference	10	3.26	3.36	.33	.63	1.95	12
Events	11	3.70	3.48	.34	.65	1.95	10
Relationships	6	1.36	2.20	.23	.48	1.45	7
Main Idea	6	1.67	1.89	.28	.25	1.49	5

Table 3C-3

Psychometric Data
Grade 2 Reading Subtests
"C" Administration
(N = 140)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Decoding Consonant Blends	12	9.83	2.49	.82	.64	1.43	67
Decoding Variant Consonants	14	8.23	4.80	.59	.71	2.49	34
Decoding Long Vowels	20	14.97	4.74	.75	.79	2.12	51
Decoding Complex Patterns	16	9.51	5.00	.59	.82	2.03	33
Compound Words	8	6.79	1.62	.85	.74	.77	82
Root Words and Affixes - D	9	6.06	1.89	.67	.59	1.14	44
Root Words and Affixes - C	8	4.72	2.62	.59	.64	1.47	42
Syllables	10	5.38	2.81	.54	.54	1.81	22
Synonyms	16	9.20	4.75	.57	.82	1.95	30
Pronoun Reference	10	4.76	3.42	.48	.67	1.85	24
Events	11	5.92	3.53	.54	.70	1.85	26
Relationships	6	2.72	2.40	.45	.60	1.39	19
Main Idea	6	2.89	2.06	.48	.44	1.40	15

Table 3C-4
Psychometric Data
Grade 2 Reading Subtests
"D" Administration
(N = 97)

Subtests	# of Items	$\bar{X}$	σ	$\bar{p}$	Alpha	Standard Error of Measurement
Compound Words	8	7.09	1.30	.88	.65	.72
Root Words and Affixes - D	9	5.92	1.84	.66	.58	1.13
Root Words and Affixes - C	8	5.35	2.48	.67	.62	1.42
Syllables	10	5.42	2.91	.54	.58	1.79
Total Comprehension	23	12.20	7.54	.53	.86	2.73

Table 3C-5
Psychometric Data
Grade 2 Mathematics Subtests
"A" Administration
(N = 150)

Subtests	No. of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent Passing 80% or more of Items in Subtest
Money	10	4.24	2.91	.42	.62	1.71	13
Linear Measurement	11	6.84	2.05	.62	.36	1.57	15
Addition No Regrouping	16	9.08	4.51	.57	.82	1.84	25
Speed Test - Addition	14	5.07	3.14	.36	.74	1.55	0
Place Value/ Numerals	25	5.50	4.63	.22	.66	2.66	0
Subtraction No Regrouping	16	6.01	4.39	.38	.76	2.09	1
Word Problems	15	3.05	2.69	.20	.45	2.13	0
Computational Transfer	22	4.05	4.62	.18	.64	2.70	0
Speed Test - Subtraction	14	4.08	3.15	.30	.66	1.76	0
Addition No Regrouping	4	2.31	1.26	.58	.41	.84	25
Subtraction No Regrouping	4	2.52	1.06	.63	.29	.78	22

Table 3C-6

Psychometric Data
Grade 2 Mathematics Subtests
"B" Administration
(N = 146)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Money	10	5.23	2.99	.52	.68	1.59	23
Linear Measurement	11	7.29	1.86	.66	.29	1.50	21
Addition No Regrouping	16	10.21	4.49	.64	.83	1.82	36
Addition with Regrouping	10	-0.26	3.04	.00	.78	1.29	2
Speed Test - Addition	14	6.67	3.72	.47	.78	1.70	11
Place Value/ Numerals	25	7.88	5.53	.32	.75	2.71	3
Subtraction no Regrouping	16	7.95	5.09	.50	.82	2.08	25
Word Problems	15	4.30	3.14	.29	.48	2.19	4
Computational Transfer	22	6.25	4.94	.28	.68	2.73	1
Speed Test - Subtraction	14	4.76	3.53	.34	.72	1.80	3
Fractions	4	1.04	1.20	.26	.00	1.05	3

Table 3C-7
Psychometric Data
Grade 2 Mathematics Subtests
"C" Administration
(N = 140)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Money	10	7.40	2.71	.74	.74	1.32	51
Linear Measurement	11	8.15	1.83	.74	.35	1.41	36
Addition No Regrouping	16	13.42	3.42	.84	.81	1.45	76
Addition with Regrouping	10	1.33	4.95	.13	.94	1.19	17
Speed Test - Addition	14	9.31	3.76	.67	.83	1.49	41
Place Value/ Numerals	25	13.01	6.40	.52	.83	2.61	13
Subtraction no Regrouping	16	12.28	4.34	.77	.85	1.64	64
Word Problems	15	6.73	4.23	.45	.74	2.09	16
Computational Transfer	22	10.13	5.65	.46	.77	2.64	07
Speed Test - Subtraction	14	7.41	4.27	.53	.83	1.69	26
Fractions	10	4.64	3.40	.46	.75	1.62	24
Subtraction with Regrouping	10	.92	4.07	.09	.88	1.34	06

Table 3C-8
Psychometric Data
Grade 2 Mathematics Subtests
"D" Administration
(N = 97)

Subtests	# of Items	$\bar{x}$	σ	$\bar{p}$	Alpha	Standard of Measurement
Place Value/ Numerals	25	14.70	6.58	.59	.85	2.51
Addition and Sub- traction No Regrouping	16	11.85	4.45	.74	.84	1.74
Speed Test Addition and Subtraction	28	16.96	6.60	.61	.88	2.27

Table 3C-9

Psychometric Data
Grade 5 Reading Subtests
"A" Administration
(N = 134)

Subtests	No. of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent Passing 80% or more of Items in Subtest
Syllables	21	8.86	6.14	.42	.77	2.84	0
Root Words and Affixes	15	4.65	3.23	.31	.49	2.24	0
Synonyms	15	5.80	4.00	.39	.69	2.13	1
Pronouns	16	7.64	4.00	.14	.67	2.24	13
Word Meaning in Context	20	6.94	5.85	.35	.80	2.57	10
Comprehension - Literal	14	5.86	4.26	.42	.73	2.10	14
Comprehension - Inferential	9	3.90	2.56	.43	.51	1.69	12
Comprehension - Main Idea - Literal	5	2.22	1.79	.44	.45	1.19	14
Comprehension - Main Idea - Inference	4	1.05	1.46	.26	.34	1.03	1
Topic Sentence	10	3.66	3.08	.37	.62	1.81	1
Fact and Opinion	10	1.86	2.71	.19	.50	1.82	0

Table 3C-10
Psychometric Data
Grade 5 Reading Subtests
"B" Administration
(N = 127)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Syllables	21	9.46	5.95	.45	.77	2.78	7
Root Words and Affixes	15	5.49	3.17	.37	.46	2.25	3
Synonyms	15	6.84	3.75	.46	.66	2.12	11
Pronouns	16	8.55	4.40	.53	.75	2.14	24
Word Meaning in Context	20	7.60	6.95	.38	.87	2.45	16
Comprehension - Literal	14	5.72	4.11	.41	.71	2.12	13
Comprehension - Inferential	9	3.95	2.76	.44	.62	1.61	14
Comprehension - Main Idea - Literal	5	2.18	1.94	.44	.56	1.16	13
Comprehension - Main Idea - Inference	4	1.32	1.43	.33	.28	1.05	5
Topic Sentence	10	3.96	3.32	.40	.70	1.72	11
Fact and Opinion	10	1.75	3.08	.17	.64	1.76	2

Table 3C-11
Psychometric Data
Grade 5 Reading Subtests
"C" Administration
(N = 123)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Syllables	21	10.62	5.56	.51	.75	2.73	10
Root Words and Affixes	15	6.56	3.21	.44	.50	2.19	03
Synonyms	15	7.94	3.73	.53	.67	2.08	18
Pronouns	16	9.69	3.82	.61	.68	2.09	28
Word Meaning in Context	20	11.05	6.86	.55	.88	2.28	36
Comprehension - Literal	14	6.93	4.40	.50	.77	2.03	24
Comprehension - Inferential	9	4.68	2.70	.52	.63	1.55	24
Comprehension - Main Idea - Literal	5	3.01	1.75	.60	.52	1.09	25
Comprehension - Main Idea - Inference	4	1.67	1.30	.42	.10	1.07	09
Topic Sentence	10	5.14	2.98	.51	.65	1.67	13
Fact and Opinion	10	3.74	3.39	.37	.71	1.74	15

Table 3C-12

Psychometric Data
Grade 5 Reading Subtests
"D" Administration
(N = 89)

Subtests	# of Items	$\bar{X}$	σ	ρ	Alpha	Standard Error of Measurement
Root Words and Affixes	15	6.67	3.83	.44	.67	2.13
Synonyms	15	8.11	3.91	.54	.72	1.99
Pronouns	16	10.34	4.43	.64	.80	1.91
Word Meaning in Context	20	10.39	7.35	.52	.90	2.27
Topic Sentence	10	4.51	3.47	.45	.76	1.63
Fact and Opinion	10	3.61	3.38	.36	.71	1.72

Table 3C-13
Psychometric Data
Grade 5 Mathematics Subtests
"A" Administration
(N = 134)

Subtests	No. of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent Passing 80% or more of Items in Subtest
Addition and Subtraction	10	7.50	2.66	.75	.73	1.31	57
Open Multiplication	4	1.68	1.63	.42	.85	.54	20
Properties and Factors	18	12.03	3.47	.67	.81	1.47	24
Number Patterns	8	2.14	2.20	.27	.44	1.53	0
Multiplication	16	7.55	5.46	.47	.87	1.90	24
Speed Test - Multiplication	14	10.66	4.05	.76	.89	1.31	62
Place Value/ Numerals	10	4.55	2.46	.46	.52	1.62	10
Word Problems	12	2.00	2.91	.17	.53	1.92	0
Perimeters	5	1.22	1.57	.24	.37	1.11	0
Area	5	.26	1.16	.05	.15	.96	0
Lines and Figures	6	.51	1.59	.09	.29	1.22	0
Number Pairs	6	.98	1.87	.16	.47	1.25	0

Table 3C-14
Psychometric Data
Grade 5 Mathematics Subtests
"B" Administration
(N = 127)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Addition and Subtraction	10	7.28	2.90	.72	.76	1.35	52
Open Multiplication	4	2.50	1.53	.63	.81	.57	38
Properties and Factors	18	12.94	3.00	.72	.78	1.36	32
Number Patterns	8	2.29	2.29	.29	.48	1.54	7
Multiplication	16	9.60	5.68	.60	.89	1.84	41
Speed Test - Multiplication	42	17.73	8.90	.42	.94	2.18	6
Place Value/ Numerals	10	5.13	2.61	.51	.58	1.60	19
Word Problems	12	1.99	3.21	.16	.62	1.88	1
Perimeters	5	1.29	1.72	.26	.50	1.09	4
Area	5	.33	1.29	.07	.26	.99	0
Lines and Figures	6	.96	1.72	.16	.36	1.26	2
Number Pairs	6	.92	1.86	.15	.47	1.23	1
Fractions	15	1.96	3.97	.13	.77	1.85	2
Division	16	4.22	5.07	.26	.82	2.09	9

Table 3C-15

Psychometric Data
Grade 5 Mathematics Subtests
"C" Administration
(N = 123)

Subtests	# of Items	$\bar{X}$	σ	$\bar{P}$	Alpha	Standard Error of Measurement	Percent passing 80% or more of Items in Subtest
Addition and Subtraction	10	7.88	2.74	.79	.76	1.26	67
Open Multiplication	4	2.85	1.19	.71	.58	.67	35
Properties and Factors	18	13.52	2.85	.75	.76	1.36	42
Number Patterns	8	3.28	2.70	.41	.65	1.50	19
Multiplication	16	11.36	4.45	.71	.83	1.77	53
Speed Test - Multiplication	42	21.62	10.48	.51	.96	2.13	15
Place Value/ Numerals	10	5.37	2.83	.54	.64	1.62	24
Word Problems	12	3.64	3.29	.30	.62	1.94	03
Perimeters	5	1.89	1.90	.38	.58	1.10	08
Area	5	.37	1.26	.07	.19	1.02	01
Lines and Figures	6	1.25	1.92	.21	.48	1.26	04
Number Pairs	6	1.69	2.18	.28	.60	1.26	08
Fractions	15	5.64	4.89	.38	.83	1.96	16
Division	16	7.33	5.57	.46	.85	2.07	25

Table 3C-16

Psychometric Data
Grade 5 Mathematics Subtests
"D" Administration
(N = 89)

Subtests	# of Items	$\bar{X}$	σ	ρ	Alpha	Standard Error of Measurement
Multiplication	16	10.75	4.90	.71	.85	1.82
Fractions	15	5.03	5.04	.38	.84	1.96
Division	16	7.79	5.78	.46	.88	1.97

Table 3C-17

Summary of Subtest Reliabilities for Grade 2 Reading

Subtest	Number of Items	Coefficients of Internal Consistency ¹								Test - Retest Reliabilities	
		Phase III-B, BTES Testing Occasions				Error Study ³ Whole Sample		Error Study Truncated Sample		Error Study Whole Sample	Error Study Truncated Sample
		A	B	C	D	Pre	Post	Pre	Post		
Decoding Consonant Blends	12	.69	.64	.64	-	.73	.74	.56	.45	.67	.34
Decoding Variant Consonants	14	.72	.73	.71	-	.79	.83	.69	.74	.73	.47
Decoding Long Vowels	20	.78	.82	.79	-	.85	.85	.69	.66	.84	.68
Decoding Complex Patterns	16	.69	.75	.82	-	.86	.89	.66	.81	.88	.78
Compound Words	8	.77	.76	.74	.65	.76	.85	.59	.78	.80	.76
Root Words and Affixes - D	9	.53	.52	.59	.58	.74	.74	.63	.70	.79	.72
Root Words and Affixes - C	8	.60	.55	.64	.62	.64	.69	.39	.54	.77	.60
Syllables	10	.36	.40	.54	.58	.65	.71	.46	.66	.73	.63
Synonyms	16	.77	.81	.82	-	.86	.88	.74	.72	.86	.73
Pronoun Reference	10	.41	.63	.67	-	.72	.77	.58	.69	.71	.53
Events ²	11	-	.65	.70		.77	.81	.72	.72	.73	.58
Relationships ²	6	-	.48	.60	.86	.65	.70	.45	.58	.68	.49
Main Idea ²	6	-	.25	.44		.63	.56	.59	.49	.57	.40

¹The reliabilities entered in the table are alpha coefficients (Cronbach, 1951).²At Occasion D the subscales labelled "Events," "Relationships," and "Main Idea" were considered one scale.³For a description of the Error Study see Wright and Kim (1977).

Table 3C-18

Summary of Subtest Reliabilities for Grade 2 Mathematics

Subtest	Number of Items	Coefficients of Internal Consistency ¹								Test - Retest Reliabilities	
		Phase III-B, BTES Testing Occasions				Error Study ² Whole Sample		Error Study Truncated Sample		Error Study Whole Sample	Error Study Truncated Sample
		A	B	C	D	Pre	Post	Pre	Post		
Money	10	.52	.68	.74	-	.74	.76	.59	.65	.75	.56
Linear Measurement	11	.36	.29	.35	-	.32	.37	.06	.39	.48	.30
Addition No Regrouping	16	.82	.83	.81	-	.86	.89	.76	.82	.72	.32
Addition With Regrouping	10	-	.78	.94	-	.94	.97	.91	.97	.78	.61
Speed Test - Addition	14	.74	.78	.83	-	.86	.89	.83	.81	.73	.52
Place Value/Numerals	25	.66	.75	.83	.85	.86	.88	.77	.83	.82	.66
Subtraction No Regrouping	16	.76	.82	.85	-	.85	.87	.76	.79	.67	.38
Word Problems	15	.45	.48	.74	-	.74	.81	.41	.68	.70	.34
Computational Transfer	22	.64	.68	.77	-	.81	.83	.61	.68	.79	.63
Addition and Subtraction No Regrouping	16	-	-	-	.84	-	-	-	-	-	-
Speed Test - Subtraction	14	.66	.72	.83	-	.87	.87	.79	.84	.64	.43
Speed Test (Total)	28	-	-	-	.88	-	-	-	-	-	-
Fractions	10	-	-	.75	-	.78	.81	.75	.78	.74	.65
Subtraction With Regrouping	10	-	-	.88	-	.87	.93	.81	.88	.77	.51

¹The reliabilities entered in the table are alpha coefficients (Cronbach, 1951).

²For a description of the Error Study see Wright and Kim (1977).

Table 3C-19

Summary of Subtest Reliabilities for Grade 5 Reading

Subtest	Number of Items	Coefficients of Internal Consistency ¹								Test - Retest Reliabilities	
		Phase III-B, BTES Testing Occasions				Error Study ² Whole Sample		Error Study Truncated Sample		Error Study Whole Sample	Error Study Truncated Sample
		A	B	C	D	Pre	Post	Pre	Post		
Syllables	21	.77	.77	.75	-	.81	.82	.68	.66	.76	.50
Root Words and Affixes	15	.49	.46	.50	.67	.70	.78	.28	.57	.70	.41
Synonyms	15	.69	.66	.67	.72	.77	.81	.39	.62	.70	.18
Pronouns	16	.67	.75	.68	.80	.79	.88	.56	.74	.78	.64
Word Meaning in Context	20	.80	.87	.88	.90	.90	.90	.74	.76	.88	.52
Comprehension - Literal	14	.73	.71	.77	-	.81	.80	.36	.57	.83	.55
Comprehension - Inferential	9	.51	.62	.63	-	.65	.70	.24	.32	.77	.56
Comprehension - Main Idea - Literal	5	.45	.56	.52	-	.60	.54	.37	.45	.60	.46
Comprehension - Main Idea - Inference	4	.34	.28	.10	-	.37	.45	.09	.05	.62	.57
Topic Sentence	10	.62	.70	.65	.76	.70	.69	.57	.66	.69	.58
Fact and Opinion	10	.50	.64	.71	.71	.71	.66	.42	.40	.70	.54

¹The reliabilities entered in the table are alpha coefficients (Cronbach, 1951).

²For a description of the Error Study see Wright and Kim (1977).

Table 3C-20

Summary of Subtest Reliabilities for Grade 5 Mathematics

Subtest	Number of Items	Coefficients of Internal Consistency ¹								Test - Retest Reliabilities	
		Phase III-B, BTES Testing Occasions				Error Study ³ Whole Sample		Error Study Truncated Sample		Error Study Whole Sample	Error Study Truncated Sample
		A	B	C	D	Pre	Post	Pre	Post		
Addition and Subtraction	10	.73	.76	.76	-	.71	.86	.61	.85	.52	.37
Open Multiplication	4	.85	.81	.58	-	.82	.67	.72	.55	.62	.24
Properties and Factors	18	.81	.78	.76	-	.83	.80	.49	.62	.65	.44
Number Patterns	8	.44	.48	.65	-	.62	.70	.50	.53	.68	.54
Multiplication	16	.87	.89	.83	.85	.91	.89	.83	.80	.68	.44
Speed Test - Multiplication ²	42	.89	.94	.96	-	.96	.96	.95	.94	.81	.67
Place Value/Numerals	10	.52	.58	.64	-	.63	.69	.46	.69	.58	.46
Word Problems	12	.53	.62	.62	-	.74	.77	.58	.70	.76	.71
Perimeter	5	.37	.50	.58	-	.67	.64	.59	.62	.58	.56
Area	5	.15	.26	.19	-	.31	.32	.34	.41	.55	.60
Lines and Figures	6	.29	.36	.48	-	.53	.54	.37	.57	.65	.67
Number Pairs	6	.47	.47	.60	-	.71	.70	.64	.63	.65	.65
Fractions	15	-	.77	.83	.84	.88	.87	.81	.85	.79	.63
Division	16	-	.82	.85	.88	.93	.91	.86	.87	.81	.64

¹The reliabilities entered in the table are alpha coefficients (Cronbach, 1951).

²The Speed Test - Multiplication had 14 items at the A testing occasion.

³For a description of the Error Study see Wright and Kim (1977).

Table 3C-21

Subtest Combinations Used in Analysis at Grade 2

READING	MATHEMATICS
1. Decoding Consonant Blends and Long Vowels (Includes: Decoding Consonant Blends, Decoding Long Vowels) (32 items) 2. Decoding Variant Consonants (14 items) 3. Decoding Complex Patterns (16 items) 4. Word Structure - Meaningful Units (Includes: Compound Words, Root Words and Affixes - D, and Root Words and Affixes - C) (25 items) 5. Syllables (10 items) 6. Total Comprehension (Includes: Events, Relationships, and Main Idea) (23 items) 7. Total Reading	1. Money (10 items) 2. Total Addition and Subtraction No Regrouping (Includes: Addition No Regrouping and Subtraction No Regrouping) (32 items) 3. Total Speed Tests (Includes: Speed Test - Addition and Speed Test - Subtraction) (28 items) 4. Place Value/Numerals (25 items) 5. Word Problems (15 items) 6. Computational Transfer (22 items) 7. Fractions (10 items) 8. Total Addition and Subtraction with Regrouping (Includes: Addition with Regrouping and Subtraction with Regrouping) (20 items) 9. Total Mathematics

Table 3C-22

Subtest Combinations Used in Analysis at Grade 5

READING	MATHEMATICS
1. Syllables (21 items) 2. Synonyms (15 items) 3. Word Meaning in Context (20 items) 4. Comprehension Main Idea, (Includes: Main Idea - Literal, Main Idea - Inference and Topic Sentence) (19 items) 5. Total Comprehension, (Includes: Comprehension - Literal, Comprehension - Inferential, Main Idea - Literal, Main Idea - Inference, Topic Sentence and Fact and Opinion) (52 items) 6. Total Reading	1. Multiplication (Includes: Multiplication and Open Multiplication) (20 items) 2. Computational Transfer (Includes: Properties and Factors and Number Patterns) (26 items) 3. Speed Test Multiplication (42 items) 4. Word Problems (12 items) 5. Total Geometry (Includes: Perimeter, Area, Lines and Figures, and Number Pairs) (22 items) 6. Fractions (15 items) 7. Division (16 items) 8. Total Mathematics

Table 3C-23

Psychometric Data on
Affect Measures
"A" Administration

Grade 2 (N=150)

<u>Affect Scale</u>	<u>Items</u>	<u>$\bar{X}$</u>	<u>σ</u>	<u>Alpha</u>	<u>Standard Error of Measurement</u>
Attitude Toward Reading	7	15.42	3.57	.71	1.77
Attitude Toward Mathematics	7	16.04	4.11	.83	1.56
Attitude Toward School	2	4.83	1.37	.58	.63

Grade 5 (N=134)

<u>Affect Scale</u>	<u>Items</u>	<u>$\bar{X}$</u>	<u>σ</u>	<u>Alpha</u>	<u>Standard Error of Measurement</u>
Attitude Toward Reading	11	42.11	7.47	.83	2.97
Attitude Toward Mathematics	11	38.70	9.88	.90	2.99
Attitude Toward School	2	7.52	2.15	.55	1.02

Table 3C-24

Psychometric Data on
Affect Measures
"B" Administration

Grade 2 (N=146)

<u>Affect Scale</u>	<u># of Items</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Alpha</u>	<u>Standard Error of Measurement</u>
Attitude Toward Reading	7	15.87	4.03	.82	1.60
Attitude Toward Mathematics	7	16.13	4.41	.88	1.44
Attitude Toward School	2	4.76	1.49	.76	.51

Grade 5 (N=127)

<u>Affect Scale</u>	<u># of Items</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Alpha</u>	<u>Standard Error of Measurement</u>
Attitude Toward Reading	11	40.47	8.67	.83	2.80
Attitude Toward Mathematics	11	36.68	10.46	.93	2.72
Attitude Toward School	2	7.43	2.21	.73	.82

Table 3C-25

Psychometric Data on
Affect Measures
"C" Administration

Grade 2 (N = 140)

<u>Affect Scale</u>	<u># of Items</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Alpha</u>	<u>Standard Error of Measurement</u>
Reading	7	16.67	3.96	.85	1.41
Mathematics	7	16.89	4.21	.89	1.28
School	2	4.97	1.23	.71	.47

Grade 5 (N = 123)

<u>Affect Scale</u>	<u># of Items</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Alpha</u>	<u>Standard Error of Measurement</u>
Reading	11	40.09	7.84	.88	2.55
Mathematics	11	37.11	10.51	.93	2.59
School	2	7.20	2.06	.72	.77

Grade two reading chart.

LOG CATEGORIES			Observation Codes
DECODING	Single consonants	1	DC
	Consonant blends & digraphs	2	DC
	Variant consonants (c or q)	3	DC
	Short vowels	4	DC
	Long vowels	5	DC
	Other decoding	6	DC
WORD STRUCTURE	Compound words	7	WS
	Root words and affixes	8	WS
	Syllables	9	WS
WORD MEANING	Synonyms	10	WM
	Pronoun reference	11	WM
	Other word meaning	12	WM
COMPREHENSION	Understanding events	13	CP
	Understanding relationships	14	CP
	Identifying main ideas	15	CP
	Other comprehension	16	CP
READING PRACTICE	Sight words (general practice at reading single words)	17	PT
	Oral reading of text	18	PT
	Silent reading of text	19	PT
	Reading in content areas	20	PT
	Listening while reading	21	PT
READING RELATED	Spelling (without specific decoding focus)	22	SP
	Grammar	23	RR
	Creative writing	24	RR
	Study skills	25	RR
	Other (copying, developmental)	26	RB,N

Grade two mathematics chart.

LOG CATEGORIES			Observation Codes
COMPUTATION	Addition/Subtraction No Regrouping - Standard Short Form	1	AS
	Addition/Subtraction No Regrouping - Instructional Algorithms with Place Value Transformations	2	PV
	Addition/Subtraction With Regrouping - Standard Short Form	3	RG
	Addition/Subtraction With Regrouping - Instructional Algorithms with Place Value Transformations	4	PV
	Speed Tests	5	AS
	Other	6	MU, AS, MR
CONCEPTS/ APPLICATION	Computational Transfer (properties, no. patterns, missing addends, number sentences with inequalities)	7	CT
	Place Value/Numerals	8	PV
	Word Problems	9	VA
	Money	10	MR
	Linear Measurement	11	SA
	Fractions	12	FR
	Developmental Activities	13	MB
	Other	14	SA, MR

Grade five reading chart.

LOG CATEGORIES			Observation Codes
WORD STRUCTURE	Root words and affixes	1	WS
	Syllables	2	WS
WORD MEANING	Synonyms	3	WM
	Pronoun reference	4	WM
	Other word meaning	5	WM
COMPREHENSION	Verbatim (no rephrasing)	6	CP
	Translation (paraphrase)	7	CP
	Inference/Synthesis	8	CP
	Identifying main ideas	9	CP
	Evaluation of fact and opinion	10	CP
	Other comprehension	11	CP
READING PRACTICE	Oral reading	12	PT
	Silent reading	13	PT
	Reading in content areas	14	PT
READING RELATED	Spelling	15	SP, DC
	Grammar	16	RR
	Creative writing	17	RR
	Study skills	18	RR
	Other	19	RR, RB, O,N

Grade five mathematics chart.

LOG CATEGORIES			Observation Codes
COMPUTATION	Addition	1	RG,AS
	Subtraction	2	RG,AS
	Multiplication - Learning basic facts	3	MU
	Speed tests in basic facts	4	MU
	Multiplication algorithm	5	MU
	Division	6	DI
	Fractions	7	FR
	Other	8	MR
CONCEPTS/ APPLICATION	Computational Transfer (factors, multiples, number patterns, properties)	9	CT
	Numerals/Place value (whole numbers)	10	PV
	Word Problems	11	VA
	Geometry Perimeter	12	SA
	Area	13	SA
	Number pairs	14	SA
	Lines or figures	15	SA
	Other	16	SA,MR

Outline of observation categories.

The following is a list of the observation categories with very brief descriptions of each.

General Content Categories

(General content includes periods other than reading or mathematics instruction. Any event that is coded in a general content category is not coded for any other categories, such as student and teacher behaviors.)

W = Wait

Wait refers to periods of no activity and no movement between activities. This would often occur when a student finishes his/her work early and no other activity is initiated.

T = Transition

Transition refers to periods of change from one activity to another. This might include lining up, taking seats, or quieting down before the next activity.

M = Management

Management refers to the conduct of class business that is unrelated to any instructional activity. This includes such things as the collection or distribution of milk money and making arrangements for a field trip.

B = Break

Break includes any recreational or free period, such as recess, unstructured P.E., lunch, milk breaks, and leaving class to use the restroom.

N = Nonacademic instruction

Nonacademic instruction refers to a wide range of organized classroom

activities that do not fall within a fairly narrow domain of academic instruction. Nonacademic instruction includes flag salutes, sharing, storytime, music, art, and structured P.E.

O = Other academic instruction

Other academic instruction refers to academic instruction other than reading and mathematics. This includes social studies and science (where there is no reading or mathematics content).

blank circle (between "O" and "DC" in the seventh column) = Absent

Reading Content Categories:

(The reading content categories correspond generally to categories found in the Reading Glossary for the daily logs at grades 2 and 5. The same observational categories are used at both grades. Therefore, some available categories will be observed primarily at one grade or the other).

DC = Decoding

WS = Word Structure

CP = Comprehension

WM = Word Meaning

PT = Practice, reading

SP = Spelling

RR = Reading Related

Reading related includes other reading content that is at or above the level of the tests to be used. Library skills would be coded RR.

RB = Reading, Below the test level

Reading below the test level includes other reading content that is at a lower level than the tests that will be used. Copying or handwriting would be coded RB.

Mathematics Content Categories:

(The mathematics content categories correspond generally to categories found in the Mathematics Glossary for the daily logs at grades 2 and 5. The same observational categories are used at both grades. Therefore, some available categories will be observed primarily at one grade or the other.)

AS = Addition and Subtraction with no regrouping, using the standard short form

RG = Regrouping, addition and subtraction using the standard short form

CT = Computational Transfer

PV = Place Value

Place value includes addition and subtraction without or with regrouping where an instructional algorithm involving a place value transformation is used.

MU = Multiplication

DI = Division

FR = Fractions

SA = Spatial Application

Spatial application includes geometry, measurement, and graphing without word problems.

VA = Verbal Application

Verbal application includes word problems.

MR = Mathematics Related

Mathematics related includes other mathematics content that is at or above the level of the tests to be used. Decimals would be coded MR.

MB = Mathematics, Below the test level

Mathematics below the test level includes other mathematics content that is at a lower level than the tests that will be used. Developmental

activities such as learning to count up to ten would be coded MB.

Setting Categories:

S = Self-paced

Self-paced includes activities where the student works at his own pace (such as seatwork).

O = Other-paced

Other-paced includes activities where the student must work at the pace of other students (such as group activities).

Difficulty Level Categories:

E = Easy

Easy includes review and practice. Few errors are made and little effort is required of the student.

M = Medium

Medium difficulty includes any activities between "easy" and "hard"

H = Hard

Hard includes those activities that the student cannot carry out. Many errors and few correct responses (about what you would expect by chance) occur.

Learner Moves:

(Learner moves coded must be those of the target student being sampled)

EW = Engaged, Written response

EO = Engaged, Oral response (statement or question)

EC = Engaged, Covert response

Engaged, Covert responses includes any student response that is generally

not observable. This includes most activities where the student is simply thinking, such as listening to the teacher or reading silently.

ED = Engaged, Directions

The above categories for engaged learner moves must involve the substantive content of the reading or mathematics coded. Engaged, directions includes any written, oral, or covert student response that involves only the directions to the reading or mathematics activity.

NI = Not engaged, Interim activity

Not engaged, interim activity refers to the nonacademic interim tasks that are part of a reading or mathematics task. This includes sharpening pencils, turning in and passing out papers, and getting books.

NW = Not engaged, Waiting for help

Not engaged, waiting for help refers to periods where the student has stopped working on a reading or mathematics task because he is waiting for help.

NO = Not engaged, Off-task

Not engaged, off-task refers to periods where the student is inappropriately disengaged from a reading or mathematics task. This would include socializing, daydreaming, and misbehavior during a reading or mathematics task.

Instructor Moves:

(An instructor move is coded only if it is received by the target student and is therefore relevant to the behavior coded for the target student at that moment. For example, if the student continues to work on a reading assignment while the teacher gives instructions to get out the mathematics text, then the student engagement in reading is coded, with no instructor move for that student.)

AO = Academic Observational monitoring

Academic observational monitoring consists of the instructor looking at or listening to a student response during a reading or mathematics task.

AF = Academic Feedback

Academic feedback refers to the teacher informing the student that a written or oral response was correct or incorrect. Any explanation is more than academic feedback.

AQ = Academic Question

An academic question consists of the teacher asking the student for a written or oral answer.

EN = Explanation, Need

Explanation, need refers to instructional explanation provided to the student to satisfy a clear and immediate student need for help.

EP = Explanation, Planned

Explanation, planned refers to instructional explanation provided according to previous plans of the instructor. This would include lectures.

Explanation, planned may involve a general student need, but not an immediate student need.

SD = Structure or Direct

As with the other instructor moves, this move must be relevant to the reading or mathematics task coded at the same moment. Structure or direct refers to statements of goals or objectives and the provision of instructions or directions.

TE = Task Engagement feedback

Task engagement feedback refers to the instructor's moves to control inappropriate student behavior, such as off task behavior, or to praise appropriate behavior, such as good work habits. However,

inappropriate and appropriate student behavior does not refer, here, to correct and incorrect responses to a reading or mathematics task. Teacher feedback for these responses would, as discussed above, be coded AF.

Source of Instructor Move:

T = Teacher

Teacher refers, here, to the instructor in charge of the class unit in which the observation takes place. This might be a reading or mathematics specialist during the student's reading or mathematics laboratory period.

O = Other adult

Other adult refers to an instructor who is not in charge of the class unit. This might be an aide or parent volunteer.

S = Student

Student is coded as the source of instruction when the target student is receiving instructional assistance (usually tutoring) from another student.

C = Curriculum

Curriculum is coded as the source of instruction when answers provided in the curriculum materials are used by the student to obtain academic feedback. Therefore, the source code, C, is used only with the instructor move code, AF.

Focus of Instructor Move:

T = Target

The target student is the focus of the instructor move only when that move is directed specifically to that individual target student.

G = Group

A group is the focus of the instructor move when the target student of the moment is in a group (two or more) that receives an instructor move. The group is coded when the instructor move is directed to the group as a whole or to some student in the group other than the target student (so long as the move is relevant to the target student).

Table 3F-1

Inter-observer Reliability Coefficients
for Reading and Mathematics Content Categories
(Grade 2, A-B Period)

	Content Category																			
	DC	WS	CP	WM	PT	SP	RR	RB	AS	RG	CT	PV	MU	DI	FR	SA	VA	MR	MS	
Reliability ¹	0.89	0.82	0.83	0.79	0.89	0.86	0.86	0.95	0.84	0.80	0.76	0.96	0.98	--	--	0.95	0.95	0.77	0.54	
Mean (Observer 1)	16.6	1.9	5.5	6.1	20.6	2.9	6.7	10.2	12.0	0.6	9.2	3.3	0.4	0.0	0.0	3.6	2.6	3.3	1.5	
Mean (Observer 2)	16.6	1.5	5.6	5.9	19.7	2.2	8.4	9.7	13.2	0.2	7.3	2.9	0.4	0.0	0.0	3.7	2.5	3.0	3.3	
Std. Dev. (Observer 1)	16.2	4.8	8.5	10.1	17.3	8.1	11.7	13.7	14.9	3.6	14.0	10.3	4.3	--	--	10.5	7.6	8.6	4.0	
Std. Dev. (Observer 2)	15.9	4.7	8.9	9.7	16.2	6.7	15.4	13.5	17.5	2.2	14.8	9.7	3.7	--	--	10.3	7.3	7.8	7.7	

¹ The coefficient is Winer's (1962) reliability for an unadjusted single measurement. It gives the reliability of the summed observation data from one observer for one student on one day.

Table 3F-2

Inter-observer Reliability Coefficients
for Reading and Mathematics Content Categories
(Grade 2, B-C Period)

	Content Category																			
	DC	WS	CP	WM	PT	SP	RR	RB	AS	RG	CT	PV	MU	DI	FR	SA	VA	MR	MS	
Reliability ¹	0.80	0.60	0.89	0.85	0.88	0.75	0.90	0.92	0.82	0.83	0.86	0.91	0.82	--	0.99	0.25	0.90	0.73	0.71	
Mean (Observer 1)	9.2	1.8	10.7	10.8	22.6	5.3	14.4	11.2	8.6	3.9	6.2	8.4	0.1	0.0	0.4	0.2	1.4	3.1	1.6	
Mean (Observer 2)	10.4	2.5	10.3	11.0	22.2	5.7	14.7	11.0	8.2	4.5	7.0	7.8	0.2	0.0	0.5	0.6	1.8	2.5	1.8	
Std. Dev. (Observer 1)	11.7	5.5	14.9	15.3	16.2	8.6	17.4	13.1	13.2	11.0	11.5	15.9	1.2	--	2.9	1.8	4.9	8.1	4.9	
Std. Dev. (Observer 2)	13.5	7.6	15.1	14.7	16.5	10.4	17.6	13.3	12.1	11.2	12.4	15.9	1.3	--	3.2	3.3	5.8	6.4	5.8	

¹ The coefficient is Winer's (1952) reliability for an unadjusted single measurement. It gives the reliability of the summed observation data from one observer for one student on one day.

Table 3F-3

Inter-observer Reliability Coefficients
for Reading and Mathematics Content Categories
(Grade 5, A-B Period)

	Content Category																			
	DC	WS	CP	WM	PT	SP	RR	RB	AS	RG	CT	PV	MU	DI	FR	SA	VA	MR	MB	
Reliability ¹	0.81	0.78	0.70	0.92	0.87	0.87	0.60	0.31	0.85	0.77	0.29	0.95	0.84	0.94	0.95	--	0.91	0.75	0.18	
Mean (Observer 1)	6.1	6.3	12.9	16.0	28.6	13.2	17.4	15.8	0.8	4.7	0.9	0.6	9.5	7.7	7.9	0.0	2.2	3.3	1.1	
Mean (Observer 2)	6.0	5.6	12.7	19.6	27.7	9.5	23.3	16.0	1.2	3.6	0.2	0.5	8.8	8.0	7.9	0.0	2.0	3.5	2.4	
Std. Dev. (Observer 1)	9.8	11.2	14.1	25.3	21.1	22.6	22.0	16.8	3.8	10.1	6.0	4.0	17.4	16.2	15.8	--	7.2	8.9	3.8	
Std. Dev. (Observer 2)	10.0	8.5	13.1	27.9	23.1	21.5	24.9	17.1	5.4	7.4	1.5	3.7	16.2	16.9	15.6	--	6.0	9.8	9.3	

¹ The coefficient is Winer's (1962) reliability for an unadjusted single measurement. It gives the reliability of the summed observation data from one observer for one student on one day.

Table 3F-4

Inter-observer Reliability Coefficients
for Reading and Mathematics Content Categories
(Grade 5, U-C Period).

	Content Category																		
	DC	WS	CP	WM	PT	SP	RR	RB	AS	RG	CT	PV	MU	DI	FR	SA	VA	MR	MB
Reliability ¹	0.53	0.82	0.88	0.82	0.88	0.82	0.64	0.58	0.07	0.20	0.92	0.19	0.93	0.94	0.98	0.95	0.79	0.92	0.84
Mean (Observer 1)	3.1	5.8	21.8	14.6	23.6	9.0	24.6	12.0	0.2	1.2	0.7	0.2	5.5	8.9	13.7	3.5	1.6	7.3	2.3
Mean (Observer 2)	4.1	5.2	23.2	15.5	21.6	9.4	26.7	9.0	0.9	0.6	0.6	0.1	6.0	9.1	13.6	2.9	2.2	7.0	2.3
Std. Dev. (Observer 1)	5.9	10.5	23.1	16.5	21.6	12.4	24.0	22.0	1.3	4.6	3.8	2.0	11.8	16.1	24.3	10.5	5.5	16.7	6.5
Std. Dev. (Observer 2)	9.4	10.9	23.7	16.8	19.7	13.0	23.2	13.3	4.7	2.7	3.8	1.1	12.8	15.5	23.1	9.0	7.8	17.0	6.8

¹ The coefficient is Winer's (1962) reliability for an unadjusted single measurement. It gives the reliability of the summed observation data from one observer for one student on one day.

Table 36-1

Correspondence Among Observation, Congruency
and Log Content CategoriesGrade 2

Observation Category	Congruency Category	Log Categories
DC	DC	Decoding, Categories 1-6
WS	WS	Word Structure, Categories 7-9
CP	CP	Comprehension, Categories 13-16
WM	WM	Word Meaning, Categories 10-12
PT	PT	Practice Reading, Categories 17-21
SP	SP	Spelling, Category 22
Sum of the Above	RM	Reading Matched, Sum of 1-22
RR+RB	RO	Reading Other, Categories 23-26
Total of DC through RB	RT	Reading Total, Sum of 1-26
AS	AS	Add/Subtract, Categories 1, 5
RG	RG	Regrouping, Category 3
CT	CT	Computational Transfer, Category 7
PV	PV	Place Value, Categories 2, 4, and 8
FR	FR	Fractions, Category 12
VA	VA	Verbal Application Category 9 (Word Problems)
Sum of the Above	MM	Math Matched, Sum of the Above
MU, DI, SA, MR, MB	MO	Math Other, Categories 6, 10, 11, 13, 14
Total of AS Thru MB	MT	Math Total, Sum of 1-14

Table 36-2

Correspondence Among Observation, Congruency
and Log Content CategoriesGrade 5

Observation Category	Congruency Category	Log Categories
DC+SP	DC/SP	Spelling, Category 15
WS	WS	Word Structure, Categories 1, 2
CP	CP	Comprehension, Categories 6-11
WM	WM	Word Meaning, Categories 3, 4, 5
PT	PT	Practice Reading, Categories 12-14
Sum of the Above	RM	Reading Matched, Sum of Above
RR&RB	RO	Reading Other, Categories 16-19
Total of DC through RB	RT	Reading Total, Categories 1-19
AS+RG	AS/RG	Add/Subtract, Categories 1, 2
CT	CT	Computational Transfer, 9
PV	PV	Place Value, Category 10
MB	MU	Multiplication, Categories 3, 4, 5
DI	DI	Division, Category 6
FR	FR	Fractions, Category 7
VA	VA	Verbal Application, Category 11 (Word Problems)
Sum of the Above	MM	Math Matched, Sum of the Above
SA, MR, MB	MO	Math Other, Categories 8, 12-16
Total of AS Through MB	MT	Math Total, Categories 1-16

Table 3G-3

Correlations between Observation
and Log Times for Content Categories
in Grade 2 Reading

Period	Content Categories								
	DC	WS	CP	WM	PT	SP	RM	RO	RT
A-B	.80	.56	.52	.46	.67	.65	.85	.65	.80
B-C	.76	.75	.67	.78	.70	.27	.79	.63	.84

Table 3G-4

Correlations between Observation
and Log Times for Content Categories
in Grade 2 Mathematics

Period	Content Categories								
	AS	RG	CT	PV	FR	VA	MM	MO	MT
A-B	.54	.75	.70	.72	--	.44	.70	.70	.71
B-C	.67	.65	.79	.61	.95	.75	.70	.55	.69

Table 3G-5

Correlations between Observation
and Log Times for Content Categories
in Grade 5 Reading

Period	Content Categories							
	DC/SP	WS	CP	WM	PT	RM	RO	RT
A-B	.32	.35	.61	.59	.42	.64	.58	.80
B-C	.30	.87	.81	.49	.71	.86	.28	.74

Table 3G-6

Correlations between Observation
and Log Times for Content Categories
in Grade 5 Mathematics

Period	Content Categories									
	AS/RG	CT	PV	MU	DI	FR	VA	MM	MO	MT
A-B	.63	.26	.85	.75	.94	.90	.43	.74	.59	.77
B-C	.76	.45	.06	.88	.89	.91	.81	.82	.87	.81

RELIABILITY OF OBSERVED ENGAGEMENT VARIABLES

The reliability of the data on engagement was determined by two quite different procedures. The first procedure examines the level of agreement between different observers coding engagement for the same students on the same day. This procedure required paired observation data. The second procedure was used with the full, unpaired observation data set so that the level of agreement between different observers and the consistency of engagement across different days (for a given student) were considered simultaneously. The latter procedure indicates how consistent engagement is across different observers and occasions for a given student, relative to the variability between student means. Intraclass correlation coefficients are used to indicate the reliability of student mean scores (across observation occasions) (Rowley, 1976).

The advantage of the first procedure is that it isolates the reliability of the coding procedures used to collect engagement data. However, the use of engagement rates, based on one day of observation per week, to adjust teacher logs that cover every school day, introduces a source of error in addition to error due to the coding procedures. This is the sampling error due to the limitation of the observation to one day per week. The second procedure for calculating reliability combines both coding error and sampling error to yield an overall reliability of projected student means for engaged time.

Table 3H-1 presents the inter-observer reliability of observed engagement as determined by the first procedure discussed above. The data are presented for grades 2 and 5 at both the A-B and B-C periods. Results are given for both total engaged time and for the proportion of engaged time to total time. The

proportion is equivalent to the engagement rate variable used to adjust the teacher logs. The engagement rate variable was also used as an outcome measure in analyses of teaching processes and student engagement. The reliability coefficient given in the second column is Winer's (1962) coefficient (r_3) for unadjusted multiple measurements. This coefficient indicates the reliability of the average of data from two observers, each coding the same student on the same day. It is "unadjusted" for any systematic differences between paired observations. The third column gives the Pearson product-moment correlation coefficient between paired observations. The column labeled "MN 1" presents the mean of the first observation in each pair. The column labeled "MN 2" presents the mean for the second observations in each pair. Similarly, "SD 1" refers to the standard deviation for the first observation in each pair, while "SD 2" refers to that for the second. "SEM" refers to the standard error of measurement. Note that the means, standard deviations, and standard errors of measurement given for total engaged time represent minutes per day.

The inter-observer reliability of the average of two observations (Winer's r_3) was used as the criterion for engagement coding procedures because the adjustment of teacher logs for engagement was based on averages of at least two observers over as many as five occasions per student within the A-B period and up to fifteen occasions per student in the B-C period. Table 3H-1 shows that this reliability was generally quite high for total engaged time, ranging from .82 to .91. It is slightly lower for the proportion of engaged time to total time, with a range of .67 to .81. This lower range of reliability for the proportion is quite acceptable for the purposes of the present study.

Table 3H-2 presents the reliability of student mean scores for observed engagement across observers and occasions, as determined by the second pro-

Table 3H-1

Inter-observer Reliability of Observed Engagement
(paired observer data)

Grade 2, A-B period (105 pairs, 210 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total engaged time	.90	.83	75.7	78.0	29.2	31.8	12.7
Proportion of engaged time to total time	.67	.50	.71	.73	.14	.15	.10

Grade 2, B-C period (239 pairs, 478 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total engaged time	.82	.70	86.5	87.9	26.8	30.7	16.1
Proportion of engaged time to total time	.70	.55	.73	.72	.15	.18	.11

Grade 5, A-B period (85 pairs, 170 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total engaged time	.91	.84	115.5	114.1	40.8	37.0	15.9
Proportion of engaged time to total time	.81	.70	.75	.73	.13	.15	.08

Grade 5, B-C period (201 pairs, 402 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total engaged time	.91	.84	115.8	120.7	41.7	41.2	16.6
Proportion of engaged time to total time	.77	.65	.72	.75	.16	.15	.09

^a Inter-observer reliability of the average of two observations, without adjustment for systematic differences (Winer, 1962).

^b Pearson product-moment correlation between paired observations.

cedure discussed previously. The data are presented for grades 2 and 5 at both the A-B and B-C periods. Results are given for engaged time in each of the following content categories: reading matched, reading total, mathematics matched, and mathematics total. The reading and mathematics "matched" categories have a direct correspondence to the aggregation of the specific content categories coded with the observation procedures (see Appendices 3D, 3E and 3G for content category definitions). The observed engagement rates applied to the logs were based upon these matched categories because of this direct correspondence in terms of specific components. The reading and mathematics "total" categories include all content coded as reading or as mathematics, including the more general reading and mathematics "related" categories and the low level reading and mathematics "below test" categories.

On the basis of the results in Table 3H-2 we concluded that the reliability of observed engagement across both observers and occasions was not sufficiently high for the A-B period to warrant adjustments of log data based on A-B period data alone. Comparing the generally high reliability for observed engagement shown for the B-C period in this table and the considerably lower coefficients for A-B engagement data, we concluded that more observations were required for a reliable estimate of mean engaged time than were afforded by the A-B period. This was attributed to the variability of student engagement across occasions. The B-C data, where approximately three times as many occasions were sampled as had been sampled for the A-B data, suggests that the low A-B reliability is due to occasion variance (or occasion by student variance). The superiority of the interobserver reliability from paired observation (Table 3H-1) where the

Table 3H-2

Reliability of Observed Engagement
Across Observers and Occasions
(complete observation data set, unpaired)

Grade	Period	N	Reliability Coefficients ^a			
			Time Engaged in Reading Matched	Time Engaged in Reading Total	Time Engaged in Mathematics Matched	Time Engaged in Mathematics Total
2	A-B	685	.51	.61	.33	.53
	B-C	2004	.73	.76	.60	.67
5	A-B	544	.44	.50	.45	.30
	B-C	1666	.76	.80	.44	.76

^aIntraclass correlation giving the reliability of student mean scores for each period (see Rowley, 1976).

variability of student engagement across occasions is not treated as error, further supports the conclusion that too few occasions in the A-B data resulted in its lower reliability.

RELIABILITY OF OBSERVED STUDENT ERROR RATES

Reliability was computed for the error rate variables using the same two procedures described previously (Appendix 3H) for the analyses of engagement. That is, the first procedure isolates the reliability of the coding procedures themselves, while the second procedure combines both coding error and sampling error to yield an overall reliability of student mean scores for several occasions.

Table 3I-1 presents the reliability results using the first procedure for observed difficulty at grade 2, while Table 3I-2 presents these results for grade 5. Information is presented for both the A-B and B-C periods for each of the following variables: total time at low error rate, the proportion of time at low error rate to total time, total time at medium error rate, the proportion of time at medium error rate to total time, total time at high error rate and the proportion of time at high error rate to total time. The reliability coefficient given in the second column is Winer's (1962) coefficient (r^3) for unadjusted multiple measurements. The third column gives the Pearson product-moment correlation coefficient between paired observations. The column labeled "MN 1" presents the mean of the first observations in each pair, while that labeled "MN 2" presents the mean for the second observations in each pair. Similarly, "SD 1" refers to the standard deviation for the first observations in each pair, while "SD2" refers to that for the second. "SEM" refers to the standard error of measurement. Note that the means, standard deviations, and standard errors of measurement given for the total time variables represent minutes per day.

Table 3I-1

Inter-observer Reliability of Observed Difficulty
(paired observer data, grade 2)

Grade 2, A-B period (105 pairs, 210 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total easy time	.77	.65	62.1	54.6	33.7	31.2	19.3
Proportion of easy time to total time	.70	.57	.60	.52	.28	.24	.17
Total medium time	.77	.64	43.5	48.7	34.0	30.5	19.6
Proportion of medium time to total time	.68	.54	.39	.45	.28	.24	.18
Total hard time	.28	.22	1.3	2.7	3.7	7.8	5.6
Proportion of hard time to total time	.38	.28	.01	.02	.04	.06	.05

Grade 2, B-C period (239 pairs, 478 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total easy time	.80	.66	66.0	64.6	34.7	38.9	21.5
Proportion of easy time to total time	.80	.67	.56	.54	.27	.29	.16
Total medium time	.83	.71	52.5	55.7	37.5	39.6	20.7
Proportion of medium time to total time	.77	.63	.42	.45	.26	.29	.17
Total hard time	.63	.52	1.6	2.3	5.7	9.4	5.7
Proportion of hard time to total time	.75	.61	.01	.02	.05	.06	.04

^a Inter-observer reliability of the average of two observations, without adjustment for systematic differences (Winer, 1962).

^b Pearson product-moment correlation between paired observations.

Table 3I-2

Inter-observer Reliability of Observed Difficulty
(paired observer data, grade 5)

Grade 5, A-B period (85 pairs, 170 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total easy time	.90	.81	77.9	78.1	50.5	51.4	22.2
Proportion of easy time to total time	.85	.74	.50	.49	.26	.28	.14
Total medium time	.85	.74	75.6	77.8	40.2	43.8	21.3
Proportion of medium time to total time	.84	.73	.49	.50	.25	.27	.14
Total hard time	.93	.88	1.6	2.4	5.8	6.9	2.3
Proportion of hard time to total time	.94	.89	.01	.02	.04	.05	.02

Grade 5, B-C period (201 pairs, 402 observations):

Variable	Reliability ^a	Pearson r^b	MN 1	MN 2	SD 1	SD 2	SEM
Total easy time	.83	.72	75.6	74.7	47.4	52.2	26.7
Proportion of easy time to total time	.75	.60	.48	.45	.26	.27	.17
Total medium time	.83	.71	81.5	82.5	49.3	48.5	26.5
Proportion of medium time to total time	.70	.54	.50	.52	.26	.27	.18
Total hard time	.66	.49	2.4	2.8	8.1	8.4	5.9
Proportion of hard time to total time	.74	.60	.02	.02	.07	.08	.05

^a Inter-observer reliability of the average of two observations, without adjustment for systematic differences (Winer, 1962).

^b Pearson product-moment correlation between paired observations.

The inter-observer reliability of the average of two observations (Winer's r_3) was used as the criterion for error rate coding procedures because the error rate variables considered for analysis involved averages of at least two observers over as many as five occasions per student for the A-B period and up to fifteen per student for the B-C period. Tables 3I-1 and 3I-2 show that this reliability coefficient for error rate is generally high, with the exception of the two high error rate variables at grade 2 in the A-B period. The fact that the reliability of the high error rate variables in the B-C period at grade 2 is substantially higher would suggest that the low grade 2 coefficients for A-B may be due to a non-representative sample of paired observations for these results. The "high" category is particularly susceptible to sampling errors because it represents a relatively rare occurrence (a mean of approximately two minutes per student per day). Given the finding that most of the paired-observer reliabilities for error rate variables were reasonably high (including the high error rate variables at grade 2 for the B-C period and grade 5 for both periods), it was concluded that the coding procedures for the error rate variables were sufficiently reliable for the purposes of this study.

Table 3I-3 presents the reliability of student mean scores for observed error rate across observers and occasions, as determined by the second procedure discussed previously. The data are presented for grades 2 and 5 at both the A-B and B-C periods for each of the three categories (low, medium, and high) within each of two content areas (reading matched and mathematics matched).

On the basis of the results in Table 3I-3, it was concluded that the reliability of observed error rate across both observers and occasions was not sufficiently high in the A-B period to warrant analyses of error rate and achievement using data from the A-B period alone. Comparing the low A-B

Table 3I-3

Reliability of Observed Difficulty
Across Observers and Occasions^a
(complete observation data set, unpaired)

Variable	Grade 2		Grade 5	
	A-B period (N = 685)	B-C period (N = 2004)	A-B period (N = 544)	B-C period (N = 1666)
Time in easy reading matched	.44	.79	.26	.71
Time in medium reading matched	.45	.78	.50	.76
Time in hard reading matched	.18	.56	-.21	.18
Time in easy mathematics matched	.39	.65	.33	.72
Time in medium mathematics matched	.27	.69	.58	.65
Time in hard mathematics matched	.65	.52	.21	.52

^a Intraclass correlation giving the reliability of student mean scores for the each period (see Rowley, 1976).

reliabilities and the considerably higher reliability for observed error rate shown for the B-C period in this table, it was concluded that more observations were required for a reliable estimate of mean error rate scores than were afforded by the A-B period. This was attributed to the variability of student error rate across occasions. The same logic applies to this conclusion as that discussed for the parallel conclusion regarding the engagement variables. That is, the greater number of occasions sampled during the B-C period accounts for that period's higher reliability. The exclusion of occasion variance from the error term for the paired observation inter-observer reliabilities (Tables 3I-1 and 3I-2) accounts for the higher reliability of these results.

Appendix 4-A

Non-Instructional Activities

Non-Instructional Activities. Among the categories presented in Table 4-1 (Chapter 4), the first four represent some form of structured learning activity, while the last four are not directly related to intended learning outcomes. The first two categories, reading/language arts and mathematics, are discussed fully in Chapter 4, where analyses are presented relating time in these areas to reading and mathematics learning outcomes. The third category, "other academic" time, included social science and physical science instruction. The fourth category, "non-academic" time, included music, art, and physical education instruction. All of these instructional activities were intended to foster some kind of learning outcome.

The last four categories in Table 4-1, however, were not intended to foster learning outcomes, at least not directly. Wait, transition, management, and break time represented intermediate functions, necessary to the instructional activities of the school day. Wait time included any time between instructional activities during which a student had no available or assigned activity. Frequently, this was a period of three or four minutes after the class had returned from a recess, during which the teacher may have organized materials or dealt with an individual student's problem. Several students might take their seats, but have no designated activity to begin. Therefore, they would simply "wait" for the teacher to initiate some activity. Transition time involved movement from one activity to another. This would include the time during which students walked into the classroom from the playground and moved to their seats. It also included time spent moving from an oral reading circle in the back of the room to individual student desks. In addition, transitions included time spent by the teacher trying to get the class to quiet down and come to order just after a transitional

movement (such as walking in from recess). Management time included any general class business, such as collecting and redistributing milk money or parental permission forms. The last non-instructional activity was break, a category that included recesses and lunch periods. To a large extent, break periods are included in the school day simply because students cannot work continuously without them. In this sense, then, break periods are intermediate to instructional activities, allowing students to relax between instructional periods.

These non-instructional activities surround and give context to ALT. However, the nature of their quantitative associations with ALT is not obvious. It is possible that time devoted to these activities simply detracted from the time available for allocation to instructional activities. However, it is also possible that these intermediate, non-instructional activities increased ALT. In particular, one might expect that longer breaks refreshed students so that they could concentrate more effectively on subsequent instruction, therefore showing higher rates of engagement.

Table 4A-1 presents intercorrelations among wait, transition, management, and break, and their correlations with allocated time, engagement rate, and residual achievement. The intercorrelations among wait, transition, management, and transition show several significant coefficients ($p \leq .10$). Where these correlations were negative, they may reasonably be attributed to ipsativity among the four variables. That is, these four variables were mutually exclusive categories, so that time coded to one of them could not be coded to any of the others. Therefore, greater amounts of time in any one category tended to reduce available time for the remaining categories. This could result in negative correlations. For example, management time was negatively correlated with transition time at both grades (significantly

Table 4A-1

Analyses of Wait, Transition, Management, and Break;
Correlations with Allocated Time^a (Reading and Mathematics), Engagement Rate
(Reading and Mathematics), and Residual Achievement (Reading and Mathematics); A-C Period

	Grade 2 (N=139) *p ≤ .10 **p ≤ .01				Grade 5 (N=122) *p ≤ .10 **p ≤ .01			
	Wait	Transition	Management	Break	Wait	Transition	Management	Break
Wait	1.00	.32**	-.35**	.26**	1.00	.08	.25**	.24**
Transition		1.00	-.40**	-.23**		1.00	-.15	.04
Management			1.00	-.24**			1.00	-.12
Break				1.00				1.00
Allocated Time, Total Reading	.02	.13	.00	-.04	-.10	.05	-.20*	-.37**
Allocated Time, Total Mathematics	.03	.23**	-.10	-.11	-.15	-.13	-.18*	.30**
Engagement Rate, Total Reading	-.41**	-.01	.09	-.21**	.04	.03	.06	-.38**
Engagement Rate, Total Mathematics	-.25**	-.05	.04	.06	.09	.03	.03	-.20*
Residual Achievement, Total Reading	-.12	.02	-.11	.01	-.09	-.20*	.04	-.09
Residual Achievement, Total Mathematics	-.09	.25**	-.07	.00	-.16*	-.07	-.10	-.23**

^aAllocated time data, used in these correlations, were obtained from observation rather than teacher logs. This procedure resulted in corresponding days of data collection for allocated time and the four variables analyzed here (wait, transition, management, and break), all of which are observation variables.

so at grade 2, but not at grade 5). It is possible that some management activities are conducted during transitions, such as passing out milk money while students are leaving the room to go to lunch. This would result in time coded to management, rather than transition. Greater amounts of such management activities would, then, be associated with less time coded to transition. Management time was also negatively associated with break time at both grades (again, significantly so only at grade 2). It may be that teachers conducted management activities just before breaks and that this management often ran over into the break period. For example, the redistribution of milk money might not be completed until after the lunch period begins, thereby cutting into lunch (break) time.

It becomes more difficult to explain these associations when the correlations occurred at one grade, but were essentially zero at the other. For example, transition time was negatively correlated with break time at grade 2 ($p \leq .10$), but correlated essentially at zero at grade 5 ($r = .04$). Long periods of transition before breaks could cut into break time, in the same way that management activities might do so (discussed above). However, it is not clear why this occurred only at grade 2, even though the average transition at grade 5 was just as long as that at grade 2 (see Table 4-1, Chapter 4). Even more puzzling is a negative correlation at one grade that is significantly positive at the other grade. This occurred for wait time and management time, which were negatively correlated at grade 2 and positively correlated at grade 5 (both significant). It is possible that the negative association at grade 2 was due to a tendency for the management activities to involve only some of the students, so that the remaining students (who had little management time) tended to wait (with no designated activity) while the teacher conducted business with the other students.

The positive association at grade 5 could be due to situations where all students were involved in the management activities, but just a few at a time. This would mean that all students would be involved in management activities part of the time, but would also wait while other students were involved in management activities at some other time. Unfortunately, there is no ready explanation for the occurrence of one pattern at grade 2 and the other at grade 5.

There was also one case of a consistent positive association. This was the association of wait and break time ($p \leq .10$ at both grades). This may have been due to the tendency for wait periods (where there was no designated activity) to occur immediately after recesses, when the teacher was busy organizing materials and/or receiving complaints arising from student behavior during the recess. More frequent breaks would thereby result in increased wait time. Wait time was also positively correlated with transition time, although significantly so only at grade 2. It is possible that some grade 2 classes were relatively inefficient, having large amounts of both transition and wait time. It is not clear, however, why this would occur at grade 2, but not grade 5.

The correlations and tentative explanations discussed above should be clearly recognized as speculative. No simple, general pattern emerges from these analyses. It is quite possible that many of these findings would not be replicated if a different sample of classes were studied, or if slightly different variable definitions were used.

In addition to the intercorrelations discussed above, Table 4A-1 presents correlations of wait, transition, management, and break with allocated time, engagement rate, and residual achievement. Correlations with allocated time are presented for both reading allocations and mathematics allocations.

One would generally expect that wait, transition, management, and break would be negatively correlated with allocated time, in that these four variables displace time that might otherwise be used for reading and mathematics allocations. Only one of eight correlations was significant at grade 2. This was a positive correlation between transition and time allocated to mathematics. Four of eight correlations were significant at grade 5, three negative and one positive. Management was negatively correlated with both reading and mathematics allocated time. Break time showed a negative correlation with allocated time in reading and a positive correlation with allocated time in mathematics.

It is most difficult to explain this combination of positive and negative correlations. As mentioned above, the negative correlations are reasonable insofar as the four non-instructional time categories displace instructional time that would otherwise be available for reading and mathematics. Explanations can be posed for the two positive correlations, but they must be considered extremely tentative. The positive association of transition time and mathematics time allocation at grade 2 could have been due to special instructional situations where additional mathematics periods were conducted in such a way that they required more frequent or longer transitions. For example, if students attend a mathematics laboratory several days each week, where the laboratory is set up in a different room than that of the regular instruction, then longer transitions would be required for movement from the regular room to the laboratory room. The nonsignificant positive correlation of transition time and reading allocated time at grade 2 could have been due to similar special instruction in reading. It is true that instructional laboratories were more common at grade 2 than at grade 5, so the lack of positive correlations between transition and allocated time at grade 5 does not contradict this explanation.

The positive association of break time and allocated time in mathematics at grade 5, juxtaposed with the negative association for break and reading time at the same grade, is quite difficult to explain. It is reasonable to believe that additional time devoted to break periods would reduce time available for instructional allocations, but it is not clear how greater time devoted to breaks would be associated with increased instructional allocations. This finding should be replicated before serious explanations are proposed. It may also be useful to investigate variables that could mediate this association of break and mathematics time.

Correlations of wait, transition, management, and break with engagement rate are presented for both reading and mathematics engagement in Table 4A-1. Five correlations were significant and all of these were negative. Wait time was negatively correlated with both reading and mathematics engagement rates at grade 2. Break time was negatively correlated with reading engagement rates at grade 2 and with both reading and mathematics engagement rates at grade 5. These results consistently disconfirm the possibility that the non-instructional activities might facilitate student engagement during the surrounding instruction. Rather, wait and break time were negatively associated with engagement, while transition and management time were unrelated with engagement. It is not clear why wait time would be negatively related with engagement only at grade 2. One possibility is that the younger, grade 2 students have less patience than their colleagues at grade 5, so that they are more easily distracted while waiting without a designated activity. This distraction could carry over into subsequent reading or mathematics periods.

The negative association of break time and reading and mathematics allocations at both grades is particularly interesting. As noted earlier, one might have anticipated the opposite finding, in that longer or more

frequent breaks might allow students to relax, refreshing their energies for greater concentration and engagement during subsequent instruction. These negative associations suggest that relatively minimal periods of break time are sufficient in terms of student engagement needs. It appears that relatively long and/or frequent breaks may establish a pattern of student "play" that carries over into periods of academic "work," resulting in lower rates of engagement.

The final group of correlations in Table 4A-1 are those of wait, transition, management, and break with reading and mathematics residual achievement. Four significant correlations appeared. Transition time was positively correlated with residual mathematics achievement at grade 2. At grade 5, transition time was negatively correlated with residual reading achievement, while wait time and break time were negatively correlated with residual mathematics achievement.

The general trend of these correlations was toward the negative. Even at grade 2, where the only significant correlation was positive, the correlations for wait time and management time with residual achievement tended to be nonsignificant negative coefficients. Both the significant and nonsignificant correlations at grade 5 were negative with only one exception ($r = .04$). The negative trend at grade 5 may be due to the disruptive characteristics of wait, transition, management, and break. These categories of behavior tend to distract students from academic concerns and often represent organizational inefficiency (as in the case of excessive wait or transition time).

It is difficult to explain the positive association of transition time and residual mathematics achievement at grade 2. As mentioned previously, it is possible that transitions are associated with special instructional

activities in grade 2 mathematics, such as learning laboratories. These special activities might, in turn, be related to higher residual achievement in mathematics. This could account for the positive association of transition time and mathematics achievement at grade 2. Note that this explanation, as many of those pertaining to Table 4A-1, is largely speculative.

In summary, the correlations presented in Table 4A-1 for the non-instructional activities (wait, transition, management, and break) pose as much puzzle as insight. Many of the findings appear to contradict each other. However, at least two patterns of merit did emerge. First, wait time and break time appeared to be negatively correlated with engagement rates. This was true for wait time at grade 2 and break time at both grades. It would appear that unstructured, non-instructional periods during the school day tend to be associated with lower levels of engagement during instruction. The other pattern was a negative association of the non-instructional activities with residual achievement at grade 5. Disruptive features of the non-instructional activities may account for this relationship.

Appendix 7-A

Correlations of All Teaching Process Measures
with Residual Achievement (A-C), by Subtest

All correlations were carried out at the student level. Class-level predictors were entered identically for each student in the class. Correlations $\geq .18$ are starred.

Grade 2 Reading

Independent Variable	Decode Blends and Long Vowels	Decode Variant Consonants	Decode Complex Patterns	Word Structure, Meaningful Units	Word Structure, Syllables	Total Comprehension	Total Reading
Planning and General Variables							
D1 Item Prediction	24*	23*	28*	17	08	13	19*
D2 Knowledge of Subject Matter	17	-18*	20*	21*	06	19*	08
D3 Differentiated Perceptions	11	-05	03	11	01	13	04
P1 Goal Orientation	-15	-14	-10	-13	-07	-05	-17
P2 Program Change Based on Need	-04	19*	01	03	07	-20*	01
P3 Flexibility of Curriculum	02	-15	-02	16	-09	07	-04
G1 Global Rating (Like Class?)	-08	12	-06	-05	-06	-16	-04
A Appropriateness of Curriculum for Individuals	02	01	-08	-01	-04	06	02
G2 Perceived Academic Competence	02	05	05	02	-05	19*	10
G3 Orientation Toward Affect	-25*	14	-21*	-15	-11	-08	-15
G4 Learning Envir. (Cooperation, St. Responsib.)	13	11	16	27*	00	38*	30*
AUN Appropriateness -Unres.	-04	09	-09	-04	-07	-06	-01
G2UN Academic Competence -Unres.	-04	12	00	-02	-08	04	05
G3UN Affect -Unres.	-24*	19*	-19*	-14	-13	-15	-13
G4UN Learning Envir. -Unres.	05	17	09	17	-04	19*	21*

Grade 2 Mathematics

Independent Variable	Add and Subtract, Without Regrouping	Add and Subtract, Speeded	Add and Subtract, With Regrouping	Computational Transfer	Place Value/ Numerals	Word Problems	Money	Fractions	Total Math
Planning and General Variables									
D1 Item Prediction	16	09	14	25*	35*	29*	15	28*	33*
D2 Knowledge of Subject Matter	01	07	-03	10	15	23*	13	-10	11
D3 Differentiated Perceptions	25*	07	-01	20*	21*	14	18*	22*	24*
P1 Goal Orientation	03	-21*	-04	00	-01	05	06	23*	-01
P2 Program Change Based on Need	01	02	11	01	03	-02	-12	05	08
P3 Flexibility of Curriculum	01	-01	-17	-04	-14	-07	03	07	-07
G1 Global Rating (Like Class?)	-03	-13	05	-03	00	03	-07	17	03
A Appropriateness of Curriculum for Individuals	21*	-01	-02	16	15	12	16	29*	21*
G2 Perceived Academic Competence	05	-03	19*	08	30*	19*	15	18*	28*
G3 Orientation Toward Affect	-16	-12	-01	-14	-04	-10	-13	02	-12
G4 Learning Envir. (Cooperation, St. Responsib.)	03	23*	20*	10	25*	19*	15	14	24*
AUN Appropriateness -Unres.	14	-09	02	10	12	11	08	33*	18*
G2UN Academic Competence -Unres.	02	-11	18*	04	23*	17	07	26*	23*
G3UN Affect -Unres.	-14	-16	02	-12	-03	-06	-14	13	-07
G4UN Learning Envir. -Unres.	00	09	18*	06	20*	17	07	22*	20*

Grade 5 Reading

Independent Variable	Total Word Meaning	Total Comprehension	Syllables	Total Reading
Planning and General Variables				
D1 Item Prediction	15	29*	05	18*
D2 Knowledge of Subject Matter	-02	06	20*	07
D3 Differentiated Perceptions	05	01	-03	05
P1 Goal Orientation	16	20*	14	22*
P2 Program Change Based on Need	06	24*	-28*	08
P3 Flexibility of Curriculum	01	12	-08	07
G1 Global Rating (Like Class?)	02	-15	03	-08
A Appropriateness of Curriculum for Individuals	17	13	12	17
G2 Perceived Academic Competence	08	13	15	10
G3 Orientation Toward Affect	-11	-14	-15	-14
G4 Learning Envir. (Cooperation, St. Responsib.)	16	11	08	15
AUN Appropriateness -Unres.	13	-01	10	06
G2UN Academic Competence -Unres.	07	-03	12	00
G3UN Affect -Unres.	-07	-19*	-10	-16
G4UN Learning Envir. -Unres.	14	00	08	07

Grade 5 Mathematics

Independent Variable	Multipli- cation	Division	Fractions	Mult. Speed Test	Total Geometry	Computa- tional Transfer	Word Problems	Total Math
D1 Item Prediction	-12	-05	00	-16	09	-07	17	-16
D2 Knowledge of Subject Matter	10	08	-10	-08	04	06	12	05
D3 Differentiated Perceptions	-12	-13	-12	-29*	-25*	01	-05	-21*
P1 Goal Orientation	-03	12	14	-01	-05	03	16	02
P2 Program Change Based on Need	20	-07	-08	-07	-06	06	-03	05
P3 Flexibility of Curriculum	12	-08	-15	-16	-19*	03	12	-09
G1 Global Rating (Like Class?)	-24*	-02	08	-04	23*	00	06	01
A Appropriateness of Curriculum for Individuals	01	12	15	-18*	-06	-01	16	-02
G2 Perceived Academic Competence	12	08	14	19*	-04	-07	01	08
G3 Orientation Toward Affect	-01	-10	-20*	-19*	-13	10	05	-06
G4 Learning Envir. (Cooperation, St. Responsib.)	02	05	07	11	-21*	-07	02	-02
AUN Appropriateness -Unres.	-14	09	17	-17	09	-01	16	-01
G2UN Academic Competence -Unres.	-11	04	15	09	15	-05	05	05
G3UN Affect -Unres.	-15	-07	-11	-18*	06	08	09	-04
G4UN Learning Envir. -Unres.	-12	03	11	07	-04	-06	05	-01

Grade 2 Reading

Independent Variable	Decode Blends and Long Vowels	Decode Variant Consonants	Decode Complex Patterns	Word Structure, Meaningful Units	Word Structure, Syllables	Total Comprehension	Total Reading
Interactive Teaching Behaviors							
Explanation-Need	06	01	07	19*	-03	-03	05
Explanation-Planned	21*	03	26*	17	-05	07	13
Academic Monitoring	09	00	06	-10	-07	-11	-05
Academic Feedback	24*	04	18*	17	-02	00	19*
Total Substantive	26*	04	22*	15	-05	-03	16
Structure/Direct	-06	-04	-00	-08	-12	03	-10
Task Engagement Feedback	-05	19*	07	-12	-10	-12	-04
Ratings Based on Interviews							
Percentage of Materials That Are Teacher-Made	10	11	16	03	00	09	08
Goal Directedness	-15	-04	-11	-09	-09	-18*	-17
Content Goal Assessment	-15	-03	-10	-15	-08	-08	-12
Specificity of Goals	-17	10	-08	-05	-08	08	-02
Awareness of Affect	-22*	18*	-18*	-14	-12	-19*	-14
Value Placed on Affect	-23*	16	-22*	-15	-06	-21*	-17
Weekly Interview Statements							
Diagnostic Statements	-28*	-03	-24*	-19*	-01	-08	-16
General Program Change Statements	04	14	09	-02	05	15	10
Teacher Attitude Questionnaire Scales							
Aspirations	-17	07	-07	-12	-05	03	-03
Attitude Toward Subject Matter	-12	-19*	-10	-06	-14	02	-11

Grade 2 Mathematics

Independent Variable	Add and Subtract, Without Regrouping	Add and Subtract, Speeded	Add and Subtract, With Regrouping	Computational Transfer	Place Value/ Numerals	Word Problems	Money	Fractions	Total Math
Interactive Teaching Behaviors									
Explanation-Need	-03	-11	-06	-18*	-13	-25*	-24*	04	-14
Explanation-Planned	01	-13	-09	12	09	07	09	07	07
Academic Monitoring	01	-04	09	09	14	16	09	-02	18*
Academic Feedback	18*	14	-02	16	23*	18*	13	-05	22*
Total Substantive	08	-03	-01	10	17	12	07	00	17
Structure/Direct	01	-08	-02	12	10	13	17	17	10
Task Engagement Feedback	12	-16	-08	-13	04	-07	-03	-14	-04
Ratings Based on Interviews									
Percentage of Materials That Are Teacher-Made	03	-13	-07	-07	-01	01	-01	01	-05
Goal Directedness	-05	-25*	-01	-05	-05	03	01	34*	00
Content Goal Assessment	07	-23*	00	01	04	08	01	20*	02
Specificity of Goals	18*	26*	06	14	15	07	15	09	17
Awareness of Affect	-06	-16	02	-11	00	-05	-08	15	00
Value Placed on Affect	-14	-18*	-04	-10	-07	-06	-16	12	-11
Weekly Interview Statements									
Diagnostic Statements	-10	-29*	02	-27*	00	-17	-13	02	-14
General Program Change Statements	01	-24*	-17	-01	-06	-01	06	04	-08
Teacher Attitude Questionnaire Scales									
Aspirations	-25*	-22*	-04	-17	-14	-03	02	16	-19*
Attitude Toward Subject Matter	-14	-01	09	-01	00	00	-16	-11	-06

Grade 5 Reading

Independent Variable	Total Word Meaning	Total Comprehension	Syllables	Total Reading
Interactive Teaching Behaviors				
Explanation-Need	05	06	-02	06
Explanation-Planned	09	-14	06	-01
Academic Monitoring	03	-05	-07	-03
Academic Feedback	14	13	07	16
Total Substantive	14	04	04	10
Structure/Direct	09	12	05	16
Task Engagement Feedback	-06	-15	-10	-10
Ratings Based on Interviews				
Percentage of Materials That Are Teacher-Made	23*	-04	09	10
Goal Directedness	13	09	11	14
Content Goal Assessment	11	00	13	07
Specificity of Goals	-13	-05	05	-08
Awareness of Affect	-07	-22*	-06	-14
Value Placed on Affect	-10	-23*	-08	-16
Weekly Interview Statements				
Diagnostic Statements	15	-17	-17	-01
General Program Change Statements	10	05	-14	05
Teacher Attitude Questionnaire Scales				
Aspirations	24*	-01	10	18*
Attitude Toward Subject Matter	03	09	08	07

Grade 5 Mathematics

Independent Variable	Multiplication	Division	Fractions	Mult. Speed Test	Total Geometry	Computational Transfer	Word Problems	Total Math
Interactive Teaching Behaviors								
Explanation-Need	06	11	03	-05	06	-12	-07	07
Explanation-Planned	-17	-07	06	-15	24*	04	-03	01
Academic Monitoring	-05	-17	-13	04	15	05	10	03
Academic Feedback	-03	08	07	03	36*	08	08	16
Total Substantive	-09	-02	02	-04	34*	04	05	11
Structure/Direct	-18*	-22*	-11	-12	-02	-09	-06	-23*
Task Engagement Feedback	-26*	-18*	-11	-17	-05	-09	00	-23*
Ratings Based on Interviews								
Percentage of Materials That Are Teacher-Made	-03	-09	20*	-01	07	-03	10	05
Goal Directedness	-18*	09	06	-04	09	10	13	01
Content Goal Assessment	-17	06	23*	-03	13	-05	16	02
Specificity of Goals	17	-21*	-10	02	-15	-07	-04	-06
Awareness of Affect	-20*	-11	04	-08	11	06	-01	02
Value Placed on Affect	-24*	-24*	-16	-10	-06	05	00	-13
Weekly Interview Statements								
Diagnostic Statements	06	-25*	03	-09	-12	-07	-03	-07
General Program Change Statements	13	-13	-16	-13	-10	03	00	-09
Teacher Attitude Questionnaire Scales								
Aspirations	-10	31*	21*	-26*	22*	08	26*	11
Attitude Toward Subject Matter	-33*	-32*	00	-07	00	-03	14	-22*

Grade 2 Reading

Independent Variable	Decode Blends and Long Vowels	Decode Variant Consonants	Decode Complex Patterns	Word Structure, Meaningful Units	Word Structure, Syllables	Total Comprehension	Total Reading
Ratings Following Observation (average across about 15 days of observation in the B-C period)							
1. Cooperation	08	22*	11	20*	00	19*	26*
2. Student Responsibility	-04	09	-02	08	-09	12	11
3. Conviviality	-16	19*	-09	-06	-09	-08	-02
4. Acceptance of Feelings	-20*	15	-12	-10	-20*	-08	-08
5. Cognitive Task Orientation	08	09	06	07	-01	02	13
6. Attention to Students	-10	28*	-03	-01	-13	-03	06
7. Expectation of Stud. Success	-03	19*	04	-01	-04	02	07
8. Knowledge of Mathematics	-06	02	-05	-11	-08	-01	-02
9. Knowledge of Reading	-03	01	00	-02	-08	08	01
10. Clarity	-10	20*	-03	-02	-06	01	07
11. Abruptness	-04	-16	-10	-11	05	-13	-18*
12. Flexibility	-29*	11	-24*	-22*	-14	-06	-15
13. Complexity of Communication	-19*	03	-17*	-13	-23*	01	-10
14. Consistency of Message	08	08	15	10	-10	09	13
15. Structuring	-08	11	00	-02	-17	04	-01

Grade 2 Mathematics

Independent Variable	Add and Subtract, Without Regrouping	Add and Subtract, Speeded	Add and Subtract, With Regrouping	Computational Transfer	Place Value/ Numerals	Word Problems	Money	Fractions	Total Math
Ratings Following Observation (average across about 15 days of observation in the B-C period)									
1. Cooperation	00	13	17	04	18*	15	09	14	17
2. Student Responsibility	-05	03	17	03	15	13	02	27*	15
3. Conviviality	-11	-16	10	-09	06	-05	-16	12	-04
4. Acceptance of Feelings	-12	-12	04	-11	-03	-03	-09	14	-04
5. Cognitive Task Orientation	09	00	19*	08	20*	16	10	35*	31*
6. Attention to Students	-03	-10	08	-01	15	09	-02	15	12
7. Expectation of Stud. Success	01	-08	19*	04	25*	11	-02	18*	20*
8. Knowledge of Mathematics	-04	-16	10	00	14	16	08	27*	16
9. Knowledge of Reading	04	-09	18*	04	27*	16	10	26*	23*
10. Clarity	-07	-11	15	02	18*	15	01	15	14
11. Abruptness	03	-03	-19*	-02	-16	-10	-08	-10	-15
12. Flexibility	-28*	-23*	03	-21*	-08	-07	-19*	-01	-23*
13. Complexity of Communication	-14	-13	11	-06	07	08	-02	20*	05
14. Consistency of Message	17	-04	22*	07	25*	16	06	16	31*
15. Structuring	11	-11	13	10	21*	19*	07	19*	26*

Grade 5 Reading

Independent Variable	Total Word Meaning	Total Comprehension	Syllables	Total Reading
Ratings Following Observation (average across about 15 days of observation in the B-C period)				
1. Cooperation	07	-03	07	00
2. Student Responsibility	23*	07	12	18*
3. Conviviality	00	-13	-11	-14
4. Acceptance of Feelings	01	-09	-13	-10
5. Cognitive Task Orientation	18*	05	15	12
6. Attention to Students	20*	03	06	10
7. Expectation of Stud. Success	06	-03	05	-01
8. Knowledge of Mathematics	04	-08	18*	-01
9. Knowledge of Reading	09	07	12	06
10. Clarity	08	02	04	03
11. Abruptness	-28*	-21*	-16	-26*
12. Flexibility	-01	-13	02	-07
13. Complexity of Communication	13	-05	11	04
14. Consistency of Message	13	06	10	10
15. Structuring	03	00	02	-02

A-12

Grade 5 Mathematics

Independent Variable	Multipli- cation	Division	Fractions	Mult. Speed Test	Total Geometry	Computa- tional Transfer	Word Problems	Total Math
Ratings Following Observation (average across about 15 days of observation in the B-C period)								
1. Cooperation	-02	12	05	04	00	02	07	08
2. Student Responsibility	-18*	-02	20*	06	-04	-13	05	-06
3. Conviviality	05	11	-07	-16	15	09	13	07
4. Acceptance of Feelings	-05	02	-11	-22*	09	07	16	-01
5. Cognitive Task Orientation	-06	03	21*	14	15	-11	04	07
6. Attention to Students	-12	03	20*	-01	15	-04	12	05
7. Expectation of Stud. Success	-18*	02	12	02	10	-05	04	-01
8. Knowledge of Mathematics	-20*	00	18*	06	17	-08	02	00
9. Knowledge of Reading	-06	04	12	08	12	-04	08	04
10. Clarity	-08	01	14	12	11	-04	04	06
11. Abruptness	-10	-08	-26*	-14	02	10	00	-07
12. Flexibility	-15	01	-06	-26*	04	06	19*	-04
13. Complexity of Communication	-14	-03	05	-02	11	-01	15	-02
14. Consistency of Message	-03	00	16	24*	05	-10	-04	07
15. Structuring	-11	01	15	12	17	-09	00	06

Appendix 7-B

Intercorrelations of Teaching Process Measures

This Appendix gives the intercorrelations of the teaching process variables used in the major analyses reported in Chapter 7. Note that D3, P3, A, G2, G3, and G4 are all residual variables with G1 partialled out. All variables are defined over the full year period. Where variables were initially measured separately for one A-B and B-C period, scores were averaged, with B-C scores given three times the weight of A-B scores. Variables G1, G2, G3, and G4 are measured generally across both reading and mathematics. All other variables are subject-matter specific. Correlations in reading are below the diagonal. Correlations in mathematics are above the diagonal.

Intercorrelations of Teaching Process Variables
(Grade 2)

	Diagnosis			Prescription			General					Interactive						
	D1	D2	D3	P1	P2	P3	A	G1	G2	G3	G4	I1	I2	I3	I4	I5	I6	I7
D1 Item Prediction	---	29	60	08	10	03	43	22	30	05	09	08	16	09	31	26	19	11
D2 Knowledge of Subject Matter	16	---	-05	-31	-03	-19	-26	46	12	-12	11	-16	00	17	29	18	14	08
D3 Differentiated Perceptions	41	27	---	39	-21	51	83	-02	31	18	-13	19	20	-01	06	14	06	10
P1 Goal Orientation	-14	-02	28	---	-45	31	64	-02	25	-05	-10	-04	15	-13	-11	-08	24	05
P2 Program Change Based on Need	-22	-39	-16	-04	---	-10	-32	20	-04	11	06	29	05	02	15	18	-20	04
P3 Flexibility of Curriculum	25	24	54	13	-18	---	51	-05	04	43	-28	09	03	-13	-31	-17	-01	09
A Appropriateness of Curriculum for Individuals	21	-04	50	40	01	13	---	-01	31	13	-11	08	16	01	-02	07	19	03
G1 Global Rating (Like Class?)	-21	-30	-01	01	56	-02	02	---	-03	-02	-01	05	-18	-14	12	-06	14	-07
G2 Perceived Academic Competence	15	20	48	25	-13	18	51	-03	---	36	35	05	40	45	14	44	12	06
G3 Orientation Toward Affect	23	-41	24	02	-05	28	33	-02	36	---	-07	24	11	19	-22	11	07	13
G4 Learning Environment (Cooperation, Student Responsibility)	-01	10	01	00	-02	20	08	-01	35	-07	---	-04	-26	-01	-13	-16	12	-28
I1 Explanation - Need	03	12	00	01	34	09	01	06	03	-08	29	---	-03	-10	-04	21	-19	13
I2 Explanation - Planned	30	27	31	06	00	16	14	-10	40	-03	01	07	---	54	26	73	13	06
I3 Academic Monitoring	03	13	20	-09	-07	-14	05	16	24	06	-31	-13	32	---	23	78	25	-05
I4 Academic Feedback	18	05	-01	-06	25	-07	07	04	03	-16	-02	16	35	11	---	65	05	18
I5 Total Substantive	21	15	11	-07	22	-06	10	08	18	-12	-08	22	56	47	91	---	14	11
I6 Structure/Direct	-02	-08	13	26	-21	02	12	04	16	26	-06	-13	07	13	-18	-10	---	-12
I7 Task Engagement Feedback	08	-15	-29	-06	21	-08	-23	13	-07	21	-19	-04	11	05	27	25	02	488

Intercorrelations of Teaching Process Variables
(Grade 5)

	Diagnosis			Prescription			General					Interactive						
	D1	D2	D3	P1	P2	P3	A	G1	G2	G3	G4	I1	I2	I3	I4	I5	I6	I7
D1 Item Prediction	---	18	-26	00	08	-14	01	36	05	-19	02	12	36	12	41	42	48	24
D2 Knowledge of Subject Matter	11	---	07	35	08	14	01	30	45	23	28	-13	24	41	26	33	27	10
D3 Differentiated Perceptions	-24	08	---	08	13	46	29	00	-06	65	05	-14	22	-20	-34	-18	03	03
P1 Goal Orientation	38	39	20	---	07	21	57	-01	65	-12	55	-10	-13	-10	-04	-13	-08	-30
P2 Program Change Based on Need	22	-09	-04	25	---	32	09	06	-14	13	-16	-09	13	09	26	19	35	-01
P3 Flexibility of Curriculum	-22	39	65	31	31	---	57	-01	19	48	27	-09	-22	-11	-36	-32	-02	07
A Appropriateness of Curriculum for Individuals	36	30	42	80	18	40	---	-01	44	-08	33	06	-03	-28	-22	-20	-29	-27
G1 Global Rating (Like Class?)	-28	30	00	-03	-16	00	-01	---	00	01	-01	-08	56	43	47	59	37	18
G2 Perceived Academic Competence	41	38	07	71	-12	04	51	00	---	-19	64	-04	-14	-09	-12	-15	-20	-37
G3 Orientation Toward Affect	-13	-03	51	-03	04	43	07	01	-19	---	-08	-09	06	17	-15	-01	17	41
G4 Learning Environment (Cooperation, Student Responsibility)	14	27	13	65	-05	31	49	01	64	-08	---	-12	-23	-29	-27	-35	-03	-11
I1 Explanation - Need	-02	-04	14	-02	20	14	-06	06	-06	15	08	---	-04	-04	07	23	-13	12
I2 Explanation - Planned	-10	01	20	-03	-18	-09	-15	28	13	13	-01	08	---	35	45	72	35	06
I3 Academic Monitoring	01	31	-36	-06	14	03	-17	56	-01	-03	14	13	25	---	45	70	27	13
I4 Academic Feedback	10	-06	-23	-06	03	-16	-33	-02	-07	-47	01	08	29	16	---	85	35	12
I5 Total Substantive	05	08	-25	-07	05	-11	-34	28	-02	-30	07	21	57	58	85	---	37	15
I6 Structure/Direct	-07	-24	-08	00	14	06	-24	-24	-12	11	04	21	16	09	33	33	---	33
I7 Task Engagement Feedback	-10	04	-01	-21	04	10	-18	11	-35	43	-12	10	08	35	-11	09	16	---

Appendix 10-A

Regressions of C Scores on A Scores, Academic Learning Time
and Teaching Process Variables on the Subsample
of Students Who Were Tested at D

Table 10-A-1

Achievement Scores at C Regressed on Achievement at A, Academic Learning Time Variables (AC), and Teaching Process Variables (AC) for Grade 2. (N=97)

Subtest	Percent of Total Variance at C Accounted for by C Scores	Percent of Residual ¹ Variance Accounted for by		
		Academic Learning Time Variables	Teaching Behavior Variables	Total
Word Structure - Meaningful Units	24	10	14	24
Word Structure - Syllables	1	19	2	21
Total Comprehension	19	15	10	26
Addition and Subtraction Speed Test	19	15	8	24
Place Value/Numerals	17	1	28	29

¹Residual variance in C is obtained by subtracting variance accounted for by pretest (A scores) from the total variance in C scores.

NOTE: These regressions parallel those reported in Chapter 7. However, these regressions were carried out on the subsample of students who were tested at D.

Table 10-A-2

Achievement Scores at C Regressed on Achievement at A, Academic Learning Time Variables (AC), and Teaching Process Variables (AC) for Grade 5. (N=89)

Subtest	Percent Total Accounted for by A Scores	Percent of Residual ¹ Variance Accounted for.		
		Academic Learning Time Variables	Teaching Behavior Variables	Total
Total Word Meaning	50	9	19	28
Multiplication	12	18	3	21
Division	19	7	9	16
Fractions	14	36	2	38

¹Residual variance in C is obtained by subtracting variance accounted for by pretest (A scores) from the total variance in C scores.

NOTE: These regressions parallel those reported in Chapter 7. However, these regressions were carried out on the subsample of students who were tested at D.